



José Francisco Andrade Albuquerque Silva

Licenciado em Ciências da Engenharia Electrotécnica e de Computadores

Classificação Automática de Imagens Aéreas Multiespectrais para Agricultura de Precisão

Dissertação para obtenção do Grau de Mestre em
Engenharia Eletrotécnica e Computadores

Orientador: Doutor José António Barata de Oliveira, Professor As-
sociado, Universidade Nova de Lisboa
Co-orientador: Francisco Marques, Engenheiro Investigador,
UNINOVA-CTS

Júri

Presidente: Doutor José Manuel Matos Ribeiro da Fonseca, FCT/UNL
Arguente: Doutor Miguel António Leão de Sousa, INIAV



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Fevereiro, 2021

Classificação Automática de Imagens Aéreas Multiespectrais para Agricultura de Precisão

Copyright © José Francisco Andrade Albuquerque Silva, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

Em primeiro lugar, quero agradecer ao meu orientador Professor José Barata, alguém que admiro por toda a sua dedicação, e pela fantástica oportunidade de trabalhar na área da robótica. Gostava também de agradecer ao meu co-orientador Francisco Marques por todo o apoio e motivação ao longo de toda a dissertação.

Um especial agradecimento aos meus companheiros de curso e tardes nos Montaditos, Filipe Antão, Luís Simões, Miguel Pato, Flávio Silva, Francisco Cerveira, Pedro Nunes e Alvaro Souza, que tornaram todo o percurso mais fácil e divertido.

Aos amigos de uma vida Filipe, Galo, Gonçalo, João, Lima, Ricardo, Tati e Veloso obrigado pelos momentos de descontração e distração, pelo apoio, pela ramboia... de uma maneira mais simples: obrigado por tornarem tudo mais fácil.

À Cristina, obrigado por tudo.

Um grande obrigado à minha família, em especial aos meus pais, pela paciência em ouvir todos os meus problemas quando as coisas corriam menos bem. Por me motivarem e fazerem de mim a pessoa que sou hoje.

RESUMO

Ao longo dos anos, a utilização dos Veículos Aéreos não Tripulados (VANT) tem vindo a aumentar quer para fins recreativos, quer para fins académicos. As grandes vantagens económicas em relação aos seus directos competidores, satélites e aeronaves, permite que esta tecnologia seja amplamente utilizada em diversas áreas de estudo, como por exemplo a agricultura de precisão. As redes neuronais convolucionais (RNCs) já provaram ser excelentes em tarefas de classificação de imagens, e quando combinadas com os VANTs, estes permitem resolver alguns dos seus principais obstáculos relacionados com a necessidade de grandes conjuntos de dados ou grande capacidade computacional. Nesta dissertação é proposto um modelo de classificação de espécies de vegetação com recurso a imagens multiespectrais de alta resolução, utilizando redes neuronais convolucionais. O conjunto de dados utilizado foi adquiridos ao longo de 3 anos em diferentes ambientes e com diferentes culturas de vegetação. Inicialmente foram realizadas algumas operações de pré processamento dos dados tais como, alinhamento e remoção da distorção. Durante a implementação foram testados diversos parâmetros de treino da RNC, e utilizados algoritmos de *Machine Learning* (ML), nomeadamente *Support Vector Machines* (SVMs), tendo sido obtidos resultados superiores a 80% de precisão.

Palavras-chave: Agricultura de Precisão, Veículos Aéreos não Tripulados, Redes Neuronais Convolucionais, Imagens Multiespectrais, Classificação de Espécies de Vegetação, *Machine Learning*, *Support Vector Machines*.

ABSTRACT

Over the years, the use of unmanned aerial vehicles (UAVs) has been increasing for both recreational and academic purposes. The great economic advantages in relation to its direct competitors, satellites and aircraft, allows this technology to be widely used in several areas of study, such as precision agriculture. Convolutional neural networks have already proven to be excellent in image classification tasks, which when combined with UAVs, allow to solve some of the main obstacles related to the need for large data sets or large computational capacity. In this dissertation, a classification model of vegetation species using high resolution multispectral images is proposed, using convolutional neural networks. The data set used was acquired over 3 years in different environments with different vegetation cultures. Initially, some data pre-processing operations were performed, such as alignment and distortion removal. During the implementation, several training parameters of the RNC were tested, and Machine Learning (ML) algorithms were used, namely Support Vector Machines (SVMs), with results above 80 % accurate.

Keywords: Precision Agriculture, Unmanned Aerial Vehicles, Convolutional Neural Networks, Multispectral Images, Classification of Vegetation Species, Machine Learning, Support Vector Machines.

ÍNDICE

Lista de Figuras	xiii
Lista de Tabelas	xvii
Siglas	xix
1 Introdução	1
1.1 Motivação	1
1.2 Objectivo e abordagem	2
1.3 Organização do Documento	3
2 Estado de Arte	5
2.1 Enquadramento Teórico	5
2.2 VANTs: Propósitos Florestais	7
2.2.1 VANTs: Vantagens e Desvantagens	7
2.2.2 Sensores Multiespectrais	10
2.2.3 VANTs e Sensores Multiespectrais: Aplicações	13
2.2.4 Discussão	15
2.3 Aplicação de Imagens Multiespectrais para Classificação de Árvores	17
2.3.1 Casos de Estudos Relacionados	17
2.3.2 Discussão	19
2.4 Métodos de <i>Machine Learning</i>	20
2.5 Redes Neurais Convolucionais	21
2.5.1 Aplicações de RNC	22
2.5.2 Fundamentos das RNC	23
2.5.3 Arquitecturas de RNCs	26
2.6 Considerações finais	28
3 Métodos e Materiais	29
3.1 Métodos	29
3.1.1 Redes Neurais Convolucionais	29

3.1.2	<i>Support Vector Machines (SVM)</i>	33
3.2	Materiais	36
3.2.1	VANT	36
3.2.2	Câmara Multiespectral	38
4	Sistema de Classificação de Vegetação	41
4.1	Apresentação dos Dados	41
4.2	Modelo Proposto : Visão Global	43
4.2.1	Ferramentas	44
4.2.2	Pré-processamento de Imagem	45
4.3	RNC: Arquitectura e Treino da rede	48
4.4	Treino do algoritmo de ML	51
4.5	Teste do algoritmo de ML	52
4.6	Visualização dos resultados	53
5	Resultados Experimentais	55
5.1	Plano de Teste	55
5.1.1	Sistema de Medição de Desempenho	55
5.1.2	Conjunto de dados de teste	56
5.1.3	Teste da RNC	57
5.1.4	Teste do algoritmo SVM	57
5.2	Avaliação e Discussão dos Resultados	58
5.2.1	Conjunto de dados	59
5.2.2	Rede Neuronal Convolucional	60
5.2.3	Algoritmo SVM	62
5.3	Resultados Gerais dos Testes	64
6	Conclusão e Trabalho Futuro	71
6.1	Conclusão	71
6.2	Trabalho Futuro	73
	Bibliografia	75

LISTA DE FIGURAS

2.1	Comparação da utilização dos VANTs, aeronaves tripuladas e satélites ao longo do tempo. Imagem adaptada de [29]	8
2.2	Comparação entre área/custo entre os VANTs, satélites e aeronaves. Figura adaptada de [29]	9
2.3	Espectros de uma câmara multiespectral com 6 canais. Figura retirada de [9]	10
2.4	Composição de uma imagem multiespectral. Figura retirada de [40] .	11
2.5	Satélites que usam sensores de aquisição de imagens multiespectrais. a) Landsat 5 equipado LTM, imagem retirada de [43]. b) IKONOS 2, imagem retirada de [22]	11
2.6	a) Área de estudo ;b) Estratégia utilizada na recolha de imagens. Imagens retiradas de [19].	14
2.7	a) VANT utilizado na recolha de imagens, b) As camaras utilizadas no estudo foram: b) uma câmara multiespectral, e c) uma câmara de imagens a cores (vermelho, verde e azul). Imagens retiradas de [33]. .	15
2.8	Planeamento da rota utilizada pelo VANT para a recolha das imagens. Imagens retiradas de [33].	15
2.9	Comparação entre algoritmos de classificação (RF e SVM). Imagem retirada de [7].	19
2.10	Arquitectura de uma rede neuronal artificial. Imagem retirada de [33].	21
2.11	Arquitectura de uma RNC, apresentando as principais camadas que a constituem.Imagem baseada em [41].	23
2.12	<i>Pooling</i> máximo e <i>Pooling</i> médio.Imagem baseada em [41].	25
2.13	Gráfico de comparação de precisão entre diversas RNCs pré-treinadas. Imagem retirada de [10].	27
3.1	Neurónio de entrada única. Imagem baseada em [16].	30
3.2	Rede Neuronal genérica. Imagem baseada em [16].	31
3.3	Convolução de uma imagem com um filtro 3x3. Imagem baseada em [42].	32

3.4	Algoritmo SVM. Ilustração do hiperplano que divide os dados de diferentes classes. Imagem retirada de [14].	33
3.5	Função de <i>Kernel</i> :quando os dados não são linearmente separáveis, são mapeados para uma dimensão superior onde podem ser separados linearmente. Imagem baseada em [49].	36
3.6	VANT utilizado (DJI Matrice 210 RTK).	37
3.7	Arquitectura do VANT.	37
3.8	Conjunto MicaSense composto pela câmara RedEdge-M, o DLS e GPS conectados entre si. Imagem retirada de [32].	39
4.1	Ilustração das diferentes lentes da câmara e a sua correspondente banda espectral. Imagem baseada em [32].	42
4.2	Esquemático do modelo de treino.	44
4.3	Exemplo de um painel de calibração de imagem.	46
4.4	a) Canal azul. b) Canal verde. c) Canal vermelho. d) Junção dos 3 canais sem alinhamento. e) Junção dos 3 canais com alinhamento.	46
4.5	Esquema representativo do pré processamento dos dados efectuado no <i>Matlab</i>	47
4.6	Esquema representativo do processo de treino da rede efectuado no <i>Matlab</i>	49
4.7	Esquemático representativo do processo de treino do algoritmo de ML efectuado no <i>Matlab</i>	52
4.8	Esquemático representativo do processo de teste do algoritmo de ML efectuado no <i>Matlab</i>	53
5.1	Alguns exemplos de falhas no alinhamento e na remoção da distorção das imagens.	59
5.2	Gráficos de comparação do treino da rede com a alteração do parâmetro das opções de treino <i>MaxEpochs</i>	60
5.3	Gráficos de comparação do treino da rede com a alteração do parâmetro das opções de treino <i>InitialLearnRate</i>	61
5.4	Exemplos dos mapas de características extraídos da RNC.	62
5.5	Representação gráfica dos resultados do processo de validação. a) Matriz Confusão. b) Curva ROC ideal.	63
5.6	Exemplo de imagens onde não está presente nenhuma classe das espécies em questão	65

5.7	Exemplo de imagens onde se verifica a grande semelhança entre espécies e a sua idêntica disposição no terreno. a)Imagem que contém videiras. b)Imagem que contém pereiras.	66
5.8	Exemplo de imagens em que não foi possível averiguar a razão para a classificação incorrecta.a)Arroz;b)Tomate;c)Tomate.	66
5.9	Matriz de confusão resultante dos testes realizados ao conjunto de dados de teste.	67
5.10	Curva ROC resultante dos testes realizados ao conjunto de dados de teste.	68

LISTA DE TABELAS

2.1	Comparação entre VANTs, Aeronaves e Satélites. Tabela adaptada de [29]	8
2.2	Características de algumas das redes pré-treinadas mais conhecidas. Tabela baseada em [24].	27
3.1	Características das bandas capturadas pela câmara RedEdge-M da MicaSense .Tabela baseada em [44].	38
3.2	Características da Câmara RedEdge-M da MicaSense. Tabela baseada em [44].	39
4.1	Análise quantitativa do conjunto de dados, bem como a sua localização no conjunto de dados.	43
4.2	Descrição das opções de treino utilizadas.	50
5.1	Descrição do conjunto de dados de teste.	57
5.2	Tabela comparativa dos três teste efectuados com os diferentes valores do parâmetro das opções de treino <i>MaxEpochs</i>	61
5.3	Tabela comparativa dos três teste efectuados com os diferentes valores do parâmetro das opções de treino <i>InitialLearnRate</i>	62
5.4	Comparação entre os melhores algoritmos de ML testados pela aplicação <i>Classification Learner</i>	63
5.5	Resultados dos testes do algoritmo SVM utilizando o conjunto de dados de validação.	64
5.6	Resultados dos testes do algoritmo SVM utilizando o conjunto de dados de teste.	65
5.7	Resultados finais do desempenho geral do sistema com base no conjunto de dados de teste.	68

SIGLAS

AVIRIS *Airborne Visible Infrared Imaging Spectrometer.*

CVH *Campo de visão Horizontal.*

DEM *Digital Elevation Model.*

DL *Deep Learning.*

DLS *Downwelling Light Sensor.*

DMS *Digital Surface Model.*

DNG *Digital Negative.*

Fn *Falso Negativo.*

Fp *Falso Positivo.*

GPS *Global Positioning System.*

IV *Infravermelho.*

JPL *Jet Propulsion Laboratory.*

k-NN *k-Nearest Neighbors.*

LTM *Landsat Thematic Mapper.*

ML *Machine Learning.*

MLP *Multilayer Perceptron.*

NDVI *Normalized Difference Vegetation Index.*

PIV	Próximo do Infravermelho.
ReLU	<i>rectified linear unit.</i>
RF	<i>Random Forest.</i>
RNA	Rede Neuronal Artificial.
RNC	Rede Neuronal Convolucional.
ROC	<i>Receiver Operating Characteristic.</i>
SANT	Sistema Aéreo não Tripulado.
SVM	<i>Support Vector Machines.</i>
TFN	Taxa Falso Negativo.
TIFF	<i>Tagged Image File Format.</i>
TVN	Taxa Verdadeiro Negativo.
TVP	Taxa Verdadeiro Positivo.
VANT	Veículo Aéreo não Tripulado.
Vn	Verdadeiro Negativo.
Vp	Verdadeiro Positivo.

INTRODUÇÃO

1.1 Motivação

Ao longo dos anos, à agricultura tem sofrido grandes alterações a todos os níveis. O grande desenvolvimento tecnológico, especialmente na área dos veículos aéreos não tripulados e na área dos sensores como câmaras multiespectrais, permitiu o desenvolvimento de sistemas autónomos e eficientes de supervisão, baixando os custos de produção e possibilitando uma produção de larga escala[2].

Os exemplos tecnológicos aplicados à agricultura são imensos, como por exemplo: análise dos solos, análise da sanidade das culturas plantadas ou desenvolvimento de mapas de produção onde é possível averiguar a presença de todas as espécies numa área de estudo. Todos os exemplos se baseiam em classificação de imagens, mais especificamente classificação de espécies de vegetação[3].

O futuro da agricultura em Portugal e no mundo passa pela agricultura de precisão, pois será a única maneira de alimentar uma população mundial crescente, provocando o menor impacto ambiental possível.

A união europeia fornece fundos aos estados membros para o desenvolvimento da sua agricultura, em particular para uma aposta na agricultura de precisão, com o intuito de todos os estados membros serem sustentáveis. No entanto um dos grandes problemas está relacionado com a fiscalização da aplicação desses fundos. O numero de beneficiários é elevado, bem como o numero de terreno a fiscalizar sendo por isso um processo muito complicado e muitas vezes impossível de executar. A utilização dos VANTs permitiu resolver alguns desses problemas, tais como o custo da recolha dos dados necessários. Novos métodos de processamento de

imagem e inteligência artificial permitem classificar imagens aéreas, auxiliando na tomada de decisão por parte do produtor, ou mesmo permite a criação de modelos de controlo totalmente autónomos [2, 3].

A classificação de espécies de vegetação tem um papel fundamental na agricultura de precisão, no entanto também pode ser igualmente importante para fins florestais.

O aumento global da temperatura da Terra está a pôr em causa a saúde e o estado das florestas, as quais representam 30% da superfície do nosso planeta. Estas constituem ecossistemas importantes para o equilíbrio do planeta Terra, sendo necessária a sua preservação, ou até mesmo a sua recuperação. Contudo é preciso compreender as florestas e para isso é necessária a recolha de grandes quantidades de dados referentes a estas [48, 50].

As aplicações existentes para a conservação e gestão sustentável das florestas, como inventários florestais, monitorização da biodiversidade, gestão de perigos e estudos de mudanças climáticas são amplamente baseadas na classificação de espécies de árvores[46].

No entanto a maioria dos métodos existentes são limitados pois a recolha de grandes quantidades de dados no terreno torna se dispendiosa uma vez que é necessário recorrer a aeronaves ou mesmo satélites. Recentemente, com o desenvolvimento da tecnologia e dada flexibilidade em adquirir dados a qualquer hora e, em qualquer lugar com logística limitada, os veículos aéreos não tripulados (VANTs) tornaram-se numa ferramenta essencial na recolha de dados florestais. Por esse motivo muitas pesquisas concentraram-se no uso de imagens provenientes de VANTs para a classificação de árvores com o objectivo de criar mapas detalhados da localização das diversas espécies permitindo desenvolver mecanismos de protecção florestal [7, 11, 35].

1.2 Objectivo e abordagem

O primeiro passo para criar esses mapas é distinguir as várias espécies de árvores e localizá-las. O objectivo principal deste trabalho é desenvolver um sistema que possa utilizar imagens aéreas como entrada e classificar o tipo de espécie de vegetação nela contida. Cada objecto possui formas, cores e texturas diferentes que permitem ao ser humano distingui-los e classificá-los. Os vários tipos de vegetação não são diferentes e cada espécie tem folhas diferentes, texturas de casca e formatos diferentes. Essas informações podem ser extraídas de imagens e usadas por técnicas de *Machine Learning* (ML) com o objectivo de classificar as diferentes

espécies.

Com os avanços recentes na área da inteligência artificial, esta começou a ser utilizada em diversos estudos relacionados com classificação de imagem tendo obtido resultados superiores aos métodos utilizados anteriormente. Estes métodos foram também aplicados à classificação espécies de vegetação, comprovando-se a capacidade das redes neuronais convolucionais (RNC) na classificação a partir de imagens recolhidas por diferentes métodos. Contudo, para o treino de uma RNC é necessária uma grande quantidade de dados, com a respectiva identificação, e uma grande capacidade de computação para permitir a identificação e classificação de imagens [6, 34, 35].

Nesta dissertação, foi utilizado um VANT equipado com um câmara multiespectral para recolher imagens aéreas com o objectivo de classificar espécies de vegetação. Para o treino e teste da RNC foram utilizadas mais de 2500 imagens aéreas de 4 espécies diferentes. Essas espécies eram videiras, arrozal, pereira e tomateiro. O conjunto de dados foi adquirido ao longo de 3 anos sendo constituído por diversos campos das várias culturas em diversos locais situados em Portugal. O modelo proposto é composto por três partes: pré-processamento, treino da rede e classificação. Na primeira fase pré-processamento ocorre um tratamento de imagem como: distorção das imagens causadas pelas lentes, alinhamento das imagens, e junção de todos os canais espectrais numa só imagem. Por fim é reajustado o tamanho das imagens de modo a facilitar todo o processamento realizado posteriormente. A fase de treino da rede é responsável desenvolver uma rede. Nesta fase ocorre a junção das imagens pré-processadas e de alguns parâmetros de treino que tornam todo o processo mais eficiente. Ainda nesta fase é necessária a estrutura de uma rede, já desenvolvida, que permita posteriormente a classificação. A fase de classificação utiliza as imagens processadas e a rede classifica as imagens recorrendo a algoritmos de ML, tais como *support vector machine* (SVM).

1.3 Organização do Documento

Este documento está estruturado em seis capítulos organizados da seguinte forma:

- **Capítulo 2 - Estado da Arte:** Neste capítulo, são apresentados apresentados alguns casos de estudo relacionados com a utilização de VANTs para fins florestais e aplicação de imagens multiespectrais para a identificação e classificação de espécies de vegetação.

- **Capítulo 3 - Métodos e Materiais:** Neste capítulo é apresentado os materiais utilizados ao longo a dissertação, bem como os métodos utilizados posteriormente na implementação do modelo e resolução do problema.
- **Capítulo 4 - Sistema de Classificação de Vegetação:** Neste capítulo é apresentado os dados disponíveis para a resolução do problema, bem como o modelo proposto nesta dissertação.
- **Capítulo 5 - Resultados Experimentais:** Neste capítulo serão apresentados os resultados obtidos, bem como a discussão dos mesmos.
- **Capítulo 6 - Conclusão:** Neste capítulo são apresentadas as considerações finais da dissertação, bem como propostas para o trabalho futuro que possa melhorar o modelo de classificação apresentado.

ESTADO DE ARTE

O presente capítulo tem como principal objectivo apresentar ao leitor os conceitos e tecnologias que podem ser aplicadas na utilização de veículos aéreos não tripulados (VANTs) para fins florestais. Inicialmente, serão abordados alguns conceitos introdutórios relacionados com a tecnologia de recolha de dados (Secção 2.1). Posteriormente serão apresentadas as vantagens e desvantagens da utilização dos VANTs para propósitos florestais, tendo em conta alguns casos de estudo da sua aplicabilidade. Serão também apresentados alguns sensores com o principal foco no sensor multiespectral (Secção 2.2). Na Secção 2.3, será finalmente abordado o tema da dissertação: Aplicação de Imagens Multiespectrais para a Classificação de Espécies de Árvores. Nas Secções seguintes são apresentados alguns métodos e algoritmos que permitem resolver o problema proposto, bem como a apresentação alguns casos semelhantes a esta dissertação (Secção 2.4 e 2.5).

2.1 Enquadramento Teórico

No mundo de hoje, a tecnologia tem cada vez mais um papel fundamental na vida humana, pois vivemos num ambiente onde a aquisição de dados é um problema, uma vez que o mundo está constantemente em evolução. Na sociedade actual as pessoas não precisam de se adaptar à tecnologia, mas esta é responsável por se modelar em torno da sociedade; processando todas as informações adquiridas por um sensor, obtendo resultados, conclusões e até mesmo aprendendo como se comportar mesmo diante de situações desconhecidas.

A tecnologia está cada vez mais presente no dia a dia, não apenas desenvolvendo-se e ajustando-se a cada indivíduo e aos seus dispositivos pessoais, mas também atingindo em grande escala diversos sectores industriais. Algumas das áreas de actuação onde a tecnologia pode dar uma grande contribuição, conseguindo o máximo de informação no menor tempo possível sobre qualquer sistema de *Big Data*, como as de vigilância de tráfego, operações de segurança e resgate, fiscalização predial, fiscalização e monitorização florestal e agricultura.

A monitorização das florestas é uma área que tem vindo a crescer gradualmente ao longo dos anos. A aplicação de diversos sensores, estrategicamente escolhidos, permite desenvolver sistemas totalmente autónomos, sendo eles bastante flexíveis e versáteis em estudos florestais. A combinação de diversos sensores, como por exemplo, sensores multiespectrais com informação aeroespacial proveniente do *Global Positioning System* (GPS) têm permitido novos desenvolvimentos como a utilização de imagens georreferenciadas que facilitam a monitorização florestal [45, 48].

Contudo, uma grande parte dos dados eram provenientes de satélites, que apesar de apresentarem grandes resolução, tinham grandes desvantagens, pois eram associados a um custo elevado e acesso limitado. No entanto, a maior desvantagem dos satélites é a própria resolução, apesar de ser muito elevada, devido à altura a que se encontra o satélite em relação ao planeta um pixel da imagem recolhida corresponde a muitos metros quadrados, impossibilitando muitas vezes obter informação útil das mesmas[39, 48].

Embora a utilização de tecnologias sensoriais seja amplamente utilizada para propósitos florestais, ainda existem grandes desafios. Uma das barreiras mais críticas às suas aplicações está relacionada com a reduzida quantidade de informação sobre as áreas de estudo. Por exemplo, como podemos verificar nos artigos [39, 48], estes estudos pretendiam avaliar surtos de pragas florestais ou propagação de incêndios. No entanto verificou-se que as imagens de satélite e fotografias aéreas de aeronaves revelaram-se inapropriadas pois nem sempre estavam disponíveis. Além disso, para a maioria das abordagens utilizadas, a resolução dos dados necessários era crítica, ou seja, os sensores utilizados necessitavam de alta resolução [51]. Devido aos grandes problemas de recolha de dados a utilização dos VANTs tem-se vindo a desenvolver sendo alvo de diversas pesquisas e estudos por parte da comunidade científica. Serão apresentados alguns casos na secção seguinte (Secção 2.2).

2.2 VANTS: Propósitos Florestais

Nesta secção serão apresentadas as vantagens e desvantagens da utilização do VANTS bem como alguns casos de estudo relacionados com eles. Será também apresentada a câmara multiespectral bem como alguns estudos onde os VANTS são equipados com este tipo de sensor.

2.2.1 VANTS: Vantagens e Desvantagens

Os recentes progressos no desenvolvimento de tecnologias de voo de dimensões reduzidas, bem como, a tendência contínua do desenvolvimento de métodos de visão computacional, permitiram desenvolver uma nova geração de plataformas autónomas, leves, de elevada resolução e com uma grande flexibilidade e versatilidade operacional [27].

Segundo o artigo [20], uma plataforma automática ou remotamente operada por uma estação de controlo, com um sistema de comunicação, incluindo um protocolo, constitui um Sistema Aéreo não Tripulado (SANT). Os SANTs são considerados sistemas completos, dos quais fazem parte: a aeronave, a estação de controlo, todos os elementos de suporte, comunicação, controlo de tráfego aéreo, lançamento e sistemas de recuperação, conforme necessário.

Os SANTs são os principais competidores no mercado contra a aquisição de dados provenientes de satélites ou aeronaves tradicionais, pois estes apresentam uma elevada acessibilidade, permitindo aceder aos dados recolhidos em tempo real. Além disso, apresentam uma resolução detalhada e um custo reduzido comparativamente às suas alternativas [29].

Os Veículos Aéreos não Tripulados (VANTS), estão incluídos na categoria de SANT, isto é, são totalmente autónomos, mas permitem serem controlados remotamente [20]. Como se pode observar na figura 2.1, a utilização de VANTS tem vindo a aumentar exponencialmente. No entanto, durante a fase de desenvolvimento e investigação dos VANTS, surgiram alguns problemas. As grandes dimensões dos sensores necessários contribuíram para o excesso de carga, que não era suportada pelo VANT, ou reduzia em muito o seu tempo de operação. Apesar da evolução constante dos VANTS estes obstáculos ainda não foram completamente superados [29].

No artigo [29] é feita uma análise comparativa entre três tipos de plataformas, um VANT, uma aeronave tradicional e um satélite, apresentada sob a forma resumida na tabela 2.1.

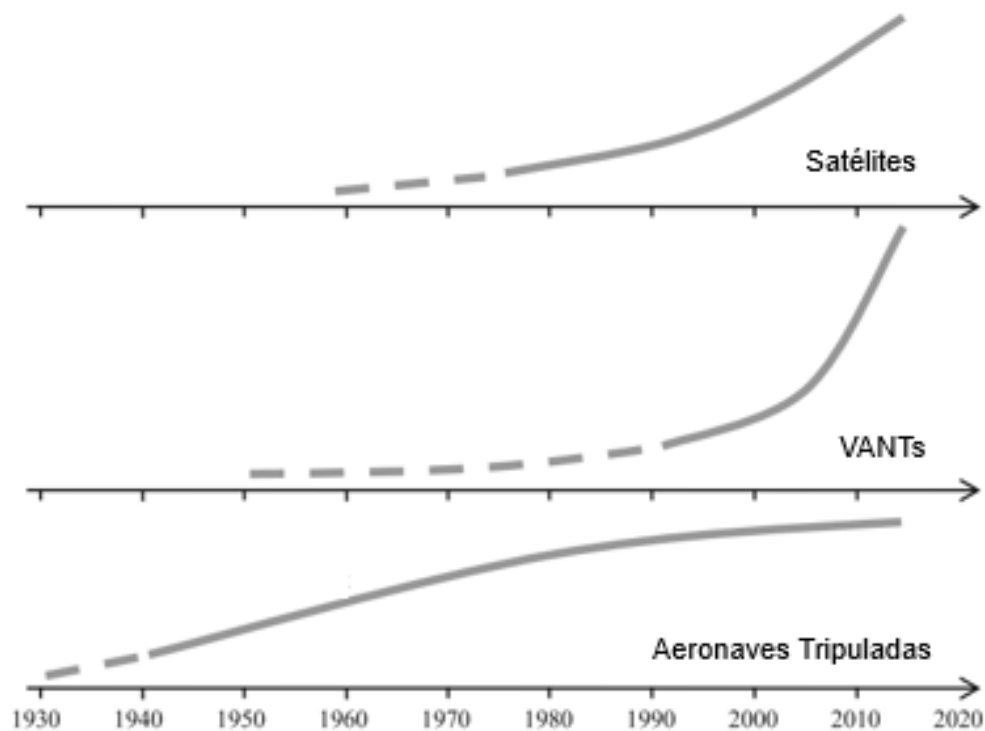


Figura 2.1: Comparação da utilização dos VANTs, aeronaves tripuladas e satélites ao longo do tempo. Imagem adaptada de [29]

		VANT	Aeronave	Satélite
Missão	Alcance	-	+	++
	Flexibilidade	++	+	-
	Autonomia	-	++	++
	Dependência da cobertura das nuvens	++	+	-
	Confiabilidade	0	+	++
Processamento	Capacidade de carga	0	+	++
	Resolução	++	+	0
	Precisão	++	+	0
	Georreferenciação	-	0	++
	Tempo de processamento	0	+	+
Legenda		(++)Muito bom	(+)Bom	(0)Médio
				(-)Mau

Tabela 2.1: Comparação entre VANTs, Aeronaves e Satélites. Tabela adaptada de [29]

A tabela 2.1 permite verificar as vantagens e desvantagens dos VANTs em relação aos seus competidores. Neste estudo, o autor integrou na tabela todos os pontos fortes e fracos das três plataformas utilizadas, verificando assim que a utilização dos VANTs apresenta grandes vantagens em relação aos seus competidores, tais como flexibilidade, resolução e precisão. Ainda é referido neste estudo, a variação do custo dos VANTs comparativamente ao das aeronaves e satélites, em relação às áreas de estudo, verificando-se que para áreas relativamente pequenas a utilização deste tipo de tecnologia permite obter uma relação custo/área muito reduzida. No entanto, com o aumento da área de estudo, essa relação torna-se menos vantajosa podendo apresentar uma relação de custo/área bastante desfavorável quando comparada com os seus competidores, como se verifica na figura 2.2.

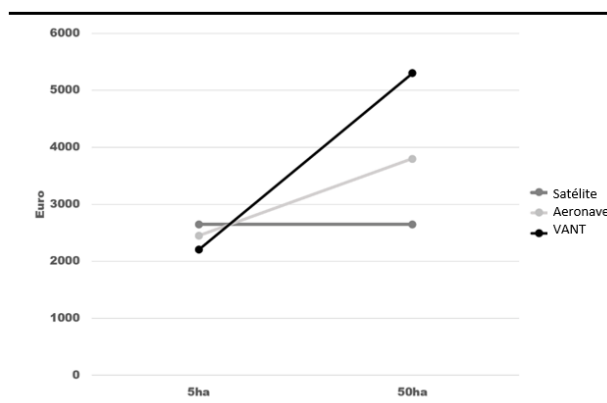


Figura 2.2: Comparação entre área/custo entre os VANTs, satélites e aeronaves. Figura adaptada de [29]

As vantagens mencionadas anteriormente permitem uma maior abrangência nas aplicações de VANTs, tais com, monitorização: em tempo real de desastres, incêndios e ainda, mapeamento, agricultura, e cada vez mais para fins florestais [29].

No artigo [8] são mencionados dois tipos de plataformas diferentes onde o autor testou as suas manobrabilidades, restrições de voo e precisão dos dados recolhidos. Ambas as plataformas foram equipadas com sensores térmicos e sensores multiespectrais capazes de recolher as bandas do espectro visível e as bandas do espectro infravermelho próximo do visível. O autor verificou que a utilização de plataformas de baixo custo permitiu a recolha de dados de alta resolução idênticos aos obtidos através de plataformas tradicionais, como aeronaves tripuladas ou até mesmo satélites. Também no artigo [17] é realizada a comparação das características de diversas plataformas VANTs, na qual é perceptível a grande

diversidade de plataformas existentes, e ainda, as suas principais funcionalidades. São também, apresentados alguns dos sensores mais utilizados, bem como a plataforma que mais se adequa a cada um deles.

Uma vez que as imagens utilizadas nesta dissertação são imagens multiespectrais, na secção seguinte será apenas abordado o sensor multiespectral e será apresentada uma breve pesquisa relacionada com a sua utilização.

2.2.2 Sensores Multiespectrais

A identificação de espécies de árvores florestais é um problema importante para o ordenamento florestal. Nos últimos anos, vários investigadores utilizaram sensores multiespectrais e hiperespectrais para identificar e classificar espécies florestais.

Inicialmente, os sensores de imagens apenas captavam a radiação visível, ou seja, os diferentes comprimentos de onda dentro do espectro visível: vermelho, verde e azul. Estes utilizavam sensores eletro-ópticos sensíveis a luz, permitindo guardar dados espectrais dentro dessa banda. Além disso, também existem sensores térmicos que, em vez de serem baseados em luz, adquirem informações com base nas emissões de temperatura corporal (ou seja, infravermelho de grande comprimento de onda).

No entanto, com o desenvolvimento de novos sensores de imagem, foi possível a utilização de novas bandas do espectro além dos domínios da luz visível, adicionando as bandas de infravermelho próximo do visível, infravermelho de curto comprimento de onda e infravermelho de longo comprimento de onda (figuras 2.4 e 2.3).

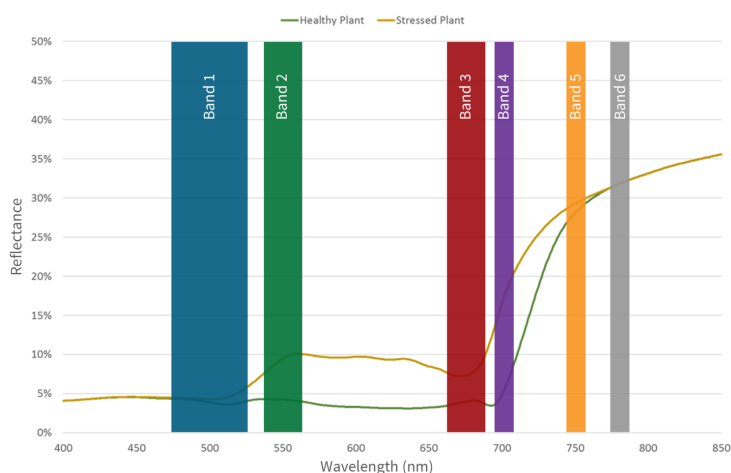


Figura 2.3: Espectros de uma câmara multiespectral com 6 canais. Figura retirada de [9]

Desta forma, a utilização das câmaras multiespectrais permitiria identificar propriedades importantes de um material ou, pelo menos, distribuir os objectos por classes baseando-se nas suas propriedades. Para aproveitar ao máximo as bandas disponíveis nas imagens multiespectrais eram realizados processamentos espectrais, também conhecidos como espectroscopia, que abrangiam a medição, a análise e a interpretação dos resultados e que revelavam informações importantes sobre as imagens em análise.

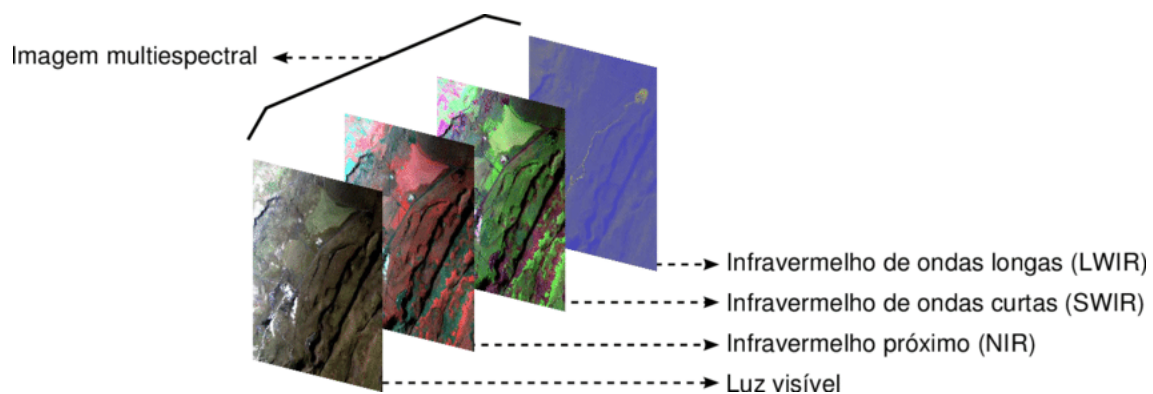


Figura 2.4: Composição de uma imagem multiespectral. Figura retirada de [40]

A utilização de sensores multiespectrais iniciou-se nos satélites, tendo vindo a desenvolver-se nos últimos 30 anos. Desde 1982, que o LANDSAT 5 equipado com o Landsat Thematic Mapper (LTM) recolhe imagens multiespectrais com uma resolução de 30 metros. A partir de 1986, o satélite SPOT XS adquiriu imagens com uma resolução de 20 metros, e posteriormente, em 1999, o IKONOS-2 fornece imagens multiespectrais com uma resolução de 4 metros [11].



Figura 2.5: Satélites que usam sensores de aquisição de imagens multiespectrais. a) Landsat 5 equipado LTM, imagem retirada de [43]. b) IKONOS 2, imagem retirada de [22]

A aplicação de sensores multiespectrais de alta resolução como o Landsat TM, SPOT XS, IKONOS e outros, apresentam grandes vantagens na identificação de

classes de espécies florestais, apesar de possuírem uma resolução espacial inferior a 60 metros [11].

O artigo [11], descreve um estudo realizado na floresta Sonian em Bruxelas onde o conjunto de imagens multiespectrais em análise foi fornecido por dois satélites IKONOS, a fim de mapear e classificar as espécies de árvores florestais. Neste exemplo de pesquisa, foram utilizadas apenas cinco das sete bandas disponíveis do domínio espectral: banda verde, banda vermelha, infravermelho próximo, infravermelho de curto comprimento de onda e infravermelho de grande comprimento de onda [11].

Além disso, o artigo [1] também menciona estudos baseados em sensores espaciais com aplicação de imagens multiespectrais na classificação de vegetação, onde além do LTM, foram estudados dados adquiridos por outros satélites, como SPOT-XS, NOAA-AVHRR e IKONOS. A partir desta pesquisa, foram obtidos resultados positivos sobre o mapeamento de espécies de vegetação com base nos índices de vegetação recolhidos.

Em 1987, as imagens multiespectrais entraram no domínio hiperespectral com o desenvolvimento do primeiro sensor de imagem aérea lançado pelo NASA *Jet Propulsion Laboratory* (JPL), designado por *Airborne Visible Infrared Imaging Spectrometer* (AVIRIS). A grande diferença entre estes dois domínios espectrais estava relacionada com número de bandas espectrais utilizadas. Enquanto o domínio multiespectral era limitado entre 3 a 10 bandas mais amplas, o domínio hiperespectral, apresentava centenas de bandas de menor dimensão, aumentando significativamente a precisão dos sensores de imagem [47].

As imagens hiperespectrais contêm, geralmente, centenas de bandas fornecendo níveis significativos de detalhes, que permitem a distinção de pequenas variações espectrais entre espécies de árvores. Apesar da abundância de informações contidas na imagem hiperespectral, discriminar espécies idênticas pode ser um desafio, muitas vezes levando a erros de classificação [47].

Segundo o artigo [13] foi descrita uma comparação entre imagens multiespectrais e hiperespectrais na classificação de espécies de árvores tropicais. As imagens hiperespectrais superaram significativamente as multiespectrais, onde se verificou que estas apresentavam algumas dificuldades em detectar pequenas variações espectrais que ocorrem entre as diferentes espécies. No entanto, as imagens hiperespectrais não apresentaram qualquer dificuldade, fornecendo informações suficientes para discriminar as diferentes espécies [5, 7, 13].

A utilização de imagens multiespectrais tem cada vez mais importância no estudo de áreas florestais. Com o desenvolvimento de novos sensores, de maior resolução e de menor dimensão, permitiu a sua utilização em VANTs. Na próxima

secção (2.2.3), serão apresentadas várias aplicações de sensores multiespectrais instalados em VANTS para fins florestais, bem como alguns casos de estudo onde foram utilizados estes tipos de tecnologias.

2.2.3 VANTS e Sensores Multiespectrais: Aplicações

Nos últimos anos, o uso de veículos aéreos não tripulados (VANTS) estendeu-se a aplicações florestais, tais como: monitorização de incêndios florestais, vigilância de desastres naturais, pesquisas científicas, monitorização de doenças florestais e, identificação e classificação de espécies de árvores florestais, entre outros [37]. Como foi referido na secção anterior, os VANTS apresentam grandes vantagens neste tipo de aplicações, como baixo custo de operação, grande capacidade de armazenamento de dados e uma grande versatilidade, permitindo adaptar a tecnologia a cada uma das aplicações.

2.2.3.1 Avaliação de parâmetros dendrométricos

Uma das aplicações de VANTS, está relacionada com a estimativa de parâmetros dimensionais das árvores florestais em estudo, como por exemplo: altura média, volume do caule, diâmetro do caule entre outros.

De acordo com o artigo [38] foi utilizado um VANT de asa fixa equipado com uma câmara S110 visível e de infravermelho. Os autores tinham como objectivo obter modelos de 3 dimensões de uma floresta situada na Noruega que permitissem estimar alguns parâmetros dendrométricos das árvores mencionadas, nomeadamente a altura média da árvore, a altura dominante, o número de caules e volume dos caules.

No artigo [19] foi utilizado um octocópter baseado num MK Okto2 modificado, equipado com uma câmara VIS-RGB Lumix G3 Panasonic com o intuito de detectar e reconstruir modelos das espécies, numa secção de floresta situada na Alemanha, com o propósito de estimar o diâmetro do caule destas.

2.2.3.2 Monitorização de incêndios e recuperações pós-incêndio

A utilização de VANTS também pode ser aplicada na monitorização de incêndios florestais ou na monitorização de processos de recuperação florestal pós incêndio.

De acordo com o artigo [4], foi utilizado um VANT eBee (senseFly Ltd., Suíça) equipado com uma câmara VIS-RGB IXUS 110 com o intuito de monitorizar os processos de recuperação pós incêndio numa floresta secundária localizada no



Figura 2.6: a) Área de estudo ;b) Estratégia utilizada na recolha de imagens. Imagens retiradas de [19].

noroeste da Itália. Os autores do presente artigo analisaram três processos de recuperação pós-incêndio distintos.

Segundo o artigo [28], o autor propôs um sistema destinado a obter automaticamente medições geométricas de incêndios florestais em tempo real, como a localização e a forma da frente de incêndio, a taxa de propagação e a altura da chama do fogo usando câmaras visuais e infravermelhas em locais terrestres fixos e câmaras montadas em VANTs. O estudo foi realizado em Portugal, entre 2001 e 2006, tendo os autores testado quatro VANTs de asa rotativa. Todos eles foram equipados com uma câmara com uma banda espectral situada na radiação visível, duas câmaras com uma banda espectral situada no infravermelho médio e duas câmaras térmicas com uma banda espectral de infravermelho térmico.

Experiências posteriores apresentaram abordagens semelhantes, no mesmo local da floresta utilizando VANTs MARVIN, Heliv e Karma, confirmando os resultados obtidos no artigo mencionado anteriormente [31].

2.2.3.3 Monitorização da Sanidade Florestal

Os VANTs também são importantes na monitorização da saúde florestal quando equipados com sensores apropriados, como sensores multiespectrais, que permitem analisar as propriedades das espécies florestais. Neste artigo [25] foi investigada a utilidade das imagens multiespectrais, adquiridas por um VANT ilustrado na figura 2.7, para produção de mapas de níveis de infestação de pragas florestais. Este estudo foi desenvolvido com o objectivo principal de apoiar decisões de intervenção no ordenamento florestal, no noroeste da Alemanha. Foi utilizado como referência o índice de vegetação com diferença normalizada modificada na classificação de imagens distinguindo-as entre cinco estados de saúde de vegetação.

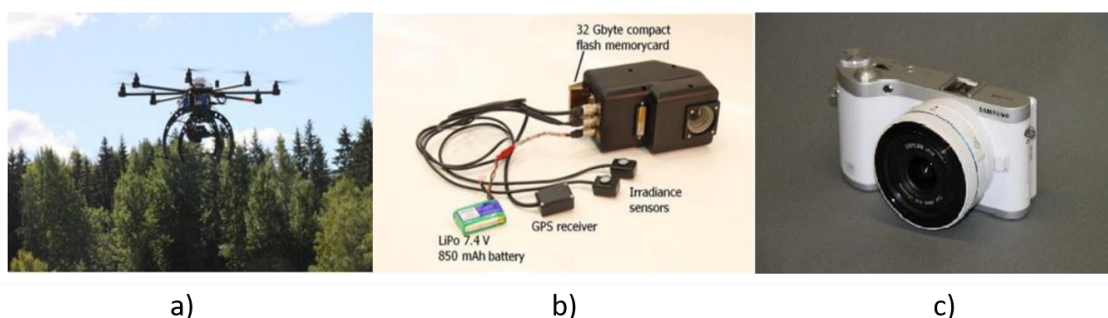


Figura 2.7: a) VANT utilizado na recolha de imagens, b) As câmaras utilizadas no estudo foram: b) uma câmara multiespectral, e c) uma câmara de imagens a cores (vermelho, verde e azul). Imagens retiradas de [33].

As imagens hiperespectrais são também, muito utilizadas. Segundo o artigo [33], o autor investigou o potencial uso de imagens hiperespectrais, adquiridas por um VANT, para mapear danos nas árvores numa floresta urbana localizada no sul da Finlândia. Um VANT equipado com uma câmara hiperespectral, configurada para recolher dados dentro de 22 bandas espectrais com o comprimento de onda entre os 500 e 900 nanómetros. Na figura 2.8 está ilustrado o percurso utilizado para a captação dos dados.

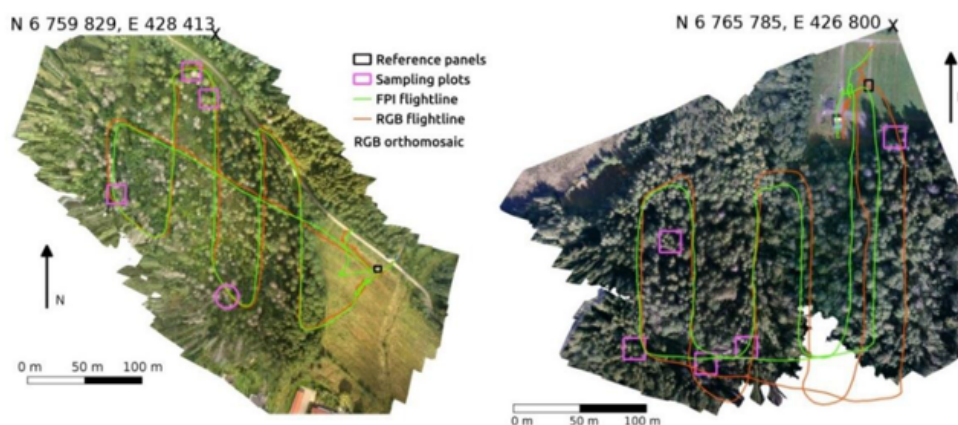


Figura 2.8: Planeamento da rota utilizada pelo VANT para a recolha das imagens. Imagens retiradas de [33].

2.2.4 Discussão

Este tipo de combinações, VANTs equipados com sensores, como câmaras multiespectrais, permitem desenvolver sistemas autónomos bastante versáteis e flexíveis para diversos cenários de operação. Pode concluir-se que a maioria dos VANTs não estão adaptados para adquirir dados de largas áreas de estudo. A capacidade

de carga não é suficiente, bem como, a dimensão dos sensores insuficientemente diminuída. Por estes factores, o tempo de voo revela-se reduzido, impossibilitando largas áreas de estudo. Para investigações florestais, a autonomia ainda é um grande desafio uma vez que, a maioria das florestas apresentam áreas muito vastas, sendo maioritariamente restringidas. No entanto, apresentam grandes vantagens como o custo da tecnologia, bem como o custo de manutenção, continuando a oferecer grande precisão e resolução nos dados recolhidos. Apresentam também, grande versatilidade e flexibilidade, permitindo adaptar os VANTs sem que isso represente custos adicionais. Permitem ainda, uma elevada aquisição de dados, sendo esta característica fundamental para a utilização de VANTs para propósitos florestais. Em relação aos sensores, estes apresentam também, vantagens e desvantagens quando comparados entre si. Como referido anteriormente, podemos concluir que a utilização de imagens hiperespectrais é mais vantajosa pois permite a recolha de dados mais pormenorizados. No entanto, a utilização de centenas de bandas espectrais introduz grande complexidade e redundância nos dados recolhidos bem como uma dificuldade acrescida a nível de computação.

O número de investigações em áreas florestais utilizando VANTs tem vindo a aumentar. Nesta secção foram apresentados alguns estudos, segundo os quais, foi possível verificar que as tecnologias sensoriais mais utilizadas são as câmaras normais, seguidas das câmaras multiespectrais. Inicialmente foram apresentados estudos relacionados com as estimativas de parâmetros dendrómetros. As avaliações destes parâmetros têm uma grande relevância no estudo florestal, uma vez que, o seu conhecimento permite calcular a massa de combustível florestal. O motivo para que esta seja uma das aplicações com maior importância está relacionado com o valor da floresta, pois este ainda é representado pela sua quantidade de madeira activa. O conhecimento dos recursos florestais está na base do processo de tomada de decisão para o planeamento florestal, como por exemplo, orientação de políticas e programas florestais. Posteriormente, foram apresentados artigos relacionados com a monitorização de incêndios e de processos de recuperação pós incêndio. Estes tipos de aplicações são extremamente importantes pois permitem criar estratégias de combate a incêndios mais adequados ao tipo de incêndio e produzir planos de recuperação florestal mais eficientes. Finalmente, foram apresentados artigos relacionados com a saúde de espécies florestais. Este tipo de aplicação tem um grande impacto ao nível da saúde florestal, uma vez que, permite a rápida intervenção no controlo de pragas ou doenças. Todas estas aplicações tem a sua relevância, pois todas contribuem para a melhoria do ordenamento florestal, a orientação de políticas e criação de estratégias que permitam manter a saúde da floresta.

2.3 Aplicação de Imagens Multiespectrais para Classificação de Árvores

A identificação e classificação de espécies de árvores florestais fornece uma ferramenta eficiente permitindo catalogar, proteger e gerir recursos florestais. Mapas florestais detalhados e precisos são cruciais para a preparação e monitorização de incêndios, secas e outros distúrbios florestais causados pelas alterações climáticas [7, 11, 36].

No entanto, a capacidade de identificar e classificar espécies é limitada, devido à falta de variação espectral, que permite discriminar pequenas diferenças espectrais entre as espécies. Na classificação das espécies, o tipo de imagem é um factor importante, uma vez que, as variáveis espaciais e resolução espectral podem influenciar a precisão da classificação [11]. A utilização de imagens multiespectrais com três a oito bandas é geralmente usada para este tipo de classificações, que varia de classificações mais gerais como o tipo de floresta a classificações mais específicas como as espécies de cada árvores [7].

2.3.1 Casos de Estudos Relacionados

Segundo o artigo [21], os autores utilizaram um VANT para adquirir imagens multiespectrais com o propósito de detectar e classificar espécies de árvores não nativas numa floresta localizada a noroeste de Itália. O objectivo do estudo era auxiliar a autoridade do parque florestal na gestão mesmo. Neste artigo, foram utilizados dois algoritmos de classificação, obtendo uma precisão geral entre os 50% e os 79%.

O artigo [11] menciona a utilização de imagens multiespectrais adquiridas por dois satélites IKONOS, com o propósito de detectar e classificar espécies de árvores florestais. O estudo foi realizado numa floresta na parte sudeste da região de Bruxelas(Bélgica) e aplicado a sete espécies de árvores diferentes de uma a duas classes etárias diferentes. Os autores recolheram imagens em diferentes épocas do ano, aproveitando as bandas espectrais mais adequadas de cada uma das épocas e, posteriormente calcularam o índice de vegetação de diferença normalizada, mais conhecido por *normalized difference vegetation index* (NDVI). Ainda, é mencionada a aplicação de um filtro de média 3 por 3 com o intuito de suavizar as imagens e aumentar a separação entre as varias espécies de árvores. Na identificação e classificação, os autores testaram as imagens recolhidas, comparando-as com uma base de dados relativa à área de estudo em questão. A precisão geral deste estudo foi, aproximadamente, 86%.

No entanto, como foi referido na secção 2.2.2, a utilização de imagens hiperespectrais na detecção e classificação das espécies de árvores apresenta melhores resultados comparativamente às imagens multiespectrais.

De acordo com o artigo [36], os autores utilizaram sensores hiperespectrais, instalados em VANTs, com o intuito de detectar e classificar espécies individuais de árvores. O estudo foi realizado numa área florestal localizada na Finlândia, tendo sido recolhidas imagens de 11 partes diferentes da área florestal em estudo. Foram utilizados pontos de controlo no solo com o objectivo de georreferenciar as imagens recolhidas. Posteriormente, os dados recolhidos foram sujeitos a um processamento rigoroso com o intuito de se extrair a informação guardada. Segundo os autores, o processamento das imagens seguiu os seguintes procedimentos: aplicação de correcções radiométricas de calibração às imagens, determinação do modelo geométrico da imagem, incluindo orientações internas e externas, elaboração de um modelo de superfície digital, *Digital Surface Model* (DSM), registo das bandas espectrais das imagens, determinação de um modelo de imagem radiométrica para transformar dados de números digitais em refletância e, cálculo dos mosaicos de imagens hiperespectrais.

Após o processamento de imagens, foram aplicados algoritmos classificadores, que permitiam através das imagens recolhidas, classificar e identificar as espécies de árvores. Os autores utilizaram os seguintes classificadores:

- *k-Nearest Neighbors* (k-NN)
- *Naive Bayes* (NB)
- *Decision Tree* (DT)
- *Multilayer Perceptron* (MLP)
- *Random Forest* (RF)

Nesse artigo são mencionadas comparações entre algoritmos de classificação descritos anteriormente, sendo que os melhores resultados foram obtidos pelos algoritmos RF e o MLP, que segundo o autor, obtiveram 95% de precisão geral. Contrariamente ao estudo anteriormente citado, o artigo [18] aplicou o o algoritmo MLP para a classificação de espécies de árvores florestais, apresentando resultados com menor precisão.

O artigo [7], apresenta a comparação entre algoritmos classificadores, que permitem a detecção e classificação de espécies florestais. Neste artigo são comparados os classificadores *support vector machines*(SVM) e *Random Forest*(RF), sendo

2.3. APLICAÇÃO DE IMAGENS MULTIESPECTRAIS PARA CLASSIFICAÇÃO DE ÁRVORES

estes aplicados a imagens multiespectrais. Verificou-se que ambos os classificadores apresentavam um valor de precisão superior a 90%.

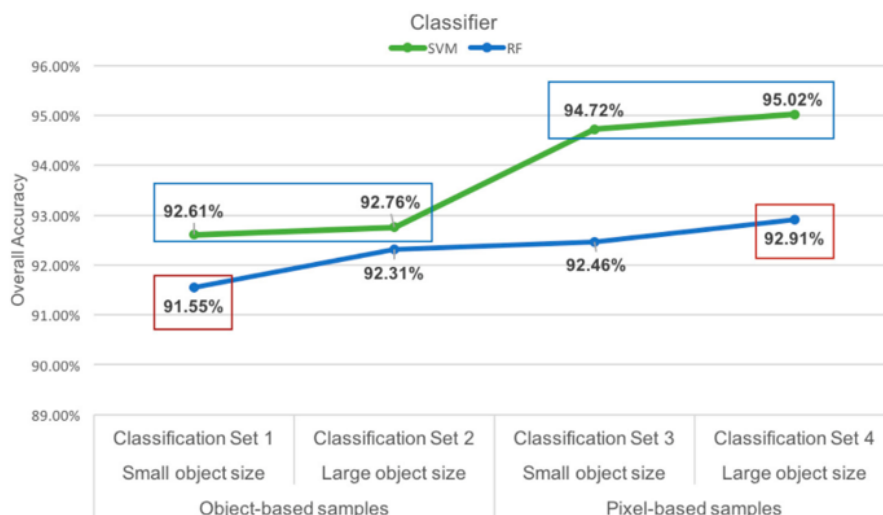


Figura 2.9: Comparação entre algoritmos de classificação (RF e SVM). Imagem retirada de [7].

2.3.2 Discussão

Nesta última secção foi apresentada uma pesquisa detalhada onde foi possível analisar diversos artigos que tinham como objectivo a classificação de espécies de árvores utilizando diferentes técnicas. Após a análise dos artigos anteriormente referidos, é possível concluir que a utilização de imagens multiespectrais permite identificar e classificar espécies de árvores. Todos os artigos apresentados, terão um papel fundamental no desenvolvimento do modelo que será proposto ao longo da dissertação. No entanto, a utilização de imagens hiperespectrais permite uma classificação mais detalhada, possibilitando obter resultados de classificação mais elevados comparativamente às imagens multiespectrais. Apesar de obter melhores resultados, as imagens hiperespectrais apresentam algumas desvantagens, nomeadamente o aumento da complexidade de todo o processamento de imagem resultante do grande número de bandas espectrais, como referido na secção 2.2.2. No entanto o sensor disponível para a realização desta investigação será a câmara multiespectral. Nos artigos apresentados, verificou-se que eram utilizadas várias técnicas de inteligência artificial, principalmente as redes neuronais convolucionais e algoritmos de *machine learning* (ML). Com base nesses artigos, verificou-se que o algoritmo de ML com os resultados de classificação neste tipo de aplicação era o SVM. Tendo em conta as principais técnicas utilizadas, serão

apresentados nas próximas secções alguns métodos de ML, bem como as redes neurais convolucionais.

2.4 Métodos de *Machine Learning*

As técnicas de Machine Learning (ML) são abordagens para a resolução de problemas de classificação e regressão, tendo a capacidade de aprenderem o funcionamento interno de um sistema a partir de um grande conjunto de dados sem conhecimentos prévios desse mesmo sistema. Têm a capacidade de modelar sistemas complexos, com uma vasta gama de variáveis de entrada sem assumir uma distribuição normal dos dados, sendo essas as principais razões pelas quais se tornaram populares. Alguns dos algoritmos ML mais comuns são *Decision Tree* (DT), Random Forest (RF), k-NN, *support vector machines* (SVM) e rede neural artificial (RNA).[7, 18, 36]

A selecção do melhor algoritmo pode ser uma tarefa desafiante, como provado pelas comparações nos artigos [18, 36]. Nestes estudos os autores testaram os mesmos algoritmos obtendo resultados totalmente distintos, quando aplicados a diversos cenários. Isto pode ser causado pelas diferentes abordagens metodológicas e características específicas dos dados, dado que o melhor algoritmo é específico para cada caso, pelo que se recomenda que a análise inclua múltiplos classificadores.

A literatura científica relativa à utilização destes métodos tem vindo a aumentar progressivamente. Os artigos [18, 36] fornecem uma visão geral dos estudos que utilizam os classificadores RF e SVM e comparam a precisão dos algoritmos na classificação de imagens. Recentemente, o interesse e a inovação têm vindo a aumentar no desenvolvimento de redes neurais profundas, uma forma de rede neuronal artificial, que supera os métodos tradicionais.

A figura 2.10 exemplifica uma arquitectura básica de rede neuronal artificial, composta por três neurónios na camada de entrada, que alimentam os neurónios numa camada oculta, e que é responsável pela activação da camada de saída. As redes neurais profundas são várias redes neuronais juntas. Vários investigadores descobriram que adicionar mais camadas à rede aumentaria a precisão do modelo. Isto tem sido comprovado em vários estudos onde redes neuronais convolucionais realizam tarefas de classificação de imagens. A secção seguinte irá focar estes modelos e a sua aplicação [6].

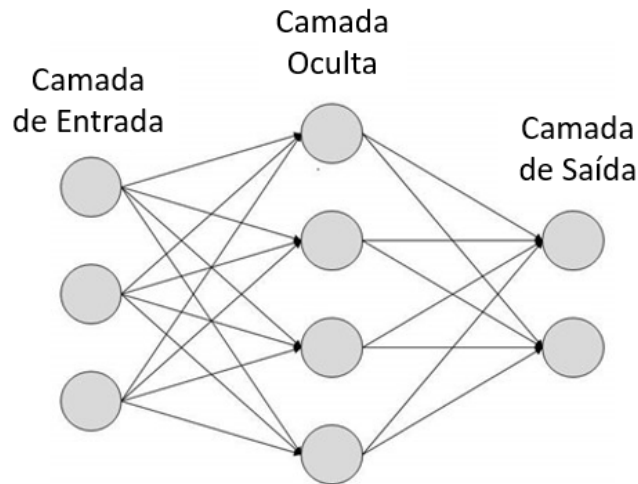


Figura 2.10: Arquitectura de uma rede neuronal artificial. Imagem retirada de [33].

2.5 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (RNCs) são uma poderosa técnica de ML. Esta estrutura de rede foi proposta pela primeira vez por Fukushima em 1988 [6]. No entanto, não foi amplamente utilizada devido aos limites de computação para o treino das redes disponíveis. Nos anos 90, alguns investigadores aplicaram um algoritmo de aprendizagem baseado em gradientes às RNCs e obtiveram resultados bem-sucedidos para o problema da classificação dos dígitos manuscritos [6]. Depois disso, vários investigadores continuaram o desenvolvimento destas redes, reportando vários estudos onde os resultados apresentados eram positivos em muitas tarefas de reconhecimento.

Em 2012, Alex Krizhevsky e colaboradores propuseram um modelo de RNC mais profundo e ganharam o desafio mais difícil para o reconhecimento visual de objectos ImageNet chamado *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC), um concurso anual no campo da visão computacional, que requer que os concorrentes desenvolvam algoritmos que possam automaticamente anotar imagens de um conjunto de dados. A *AlexNet* alcançou uma precisão de reconhecimento de última geração contra todas as abordagens tradicionais de ML e de visão computacional. Os autores explicaram que as limitações no seu modelo se deviam a restrições tecnológicas, bem como à quantidade de dados disponíveis, e previram que com mais dados e avanços tecnológicos este tipo de modelos aumentaria a precisão[6]. Foi um avanço significativo no campo da ML e da visão computacional para tarefas de reconhecimento visual e classificação. É também, a altura na história da inteligência artificial em que o interesse nas RNC aumentou

exponencialmente.

Com o aumento da computação foi possível começar a classificar as imagens, individualmente, ao nível dos seus pixels. O ano de 2015 foi outra importante referência na segmentação semântica, tendo-se alcançado o objectivo de atribuir cada pixel de uma imagem a uma classe. Paralelamente ao ajustamento feito à arquitectura e precisão da RNC, foram tomadas outras medidas para tornar o processo mais rápido e acrescentar características, tais como a definição dos limites dos objectos. Este método combinou a classificação e detecção da imagem e iniciou uma nova tendência no terreno [6].

2.5.1 Aplicações de RNC

Com o desenvolvimento das RNC, estas tornaram-se populares e alvo de vários estudos, muitos deles com o objectivo de classificar de imagens. Os autores do artigo [34], utilizaram as RNCs com a finalidade de classificar algumas espécies florestais através de imagens a cores. Essas imagens foram recolhidas por um VANT equipado com um câmara de alta definição. O artigo anterior comprovou que as RNCs permitem obter bons resultados em classificação de imagens, nomeadamente imagens de espécies de árvores.

O artigo [23], propõe um método de classificação de imagens baseado em RNC com o intuito de classificar diversos tipos de ervas daninha. Neste artigo foi utilizado um VANT com sensores de imagem de baixo custo sendo que o objectivo principal era desenvolver um modelo optimizado de classificação, que através dos dados adquiridos com um custo reduzido, permitisse distinguir diversos tipos espécies infestantes. Os autores reportam resultados com valores de precisão superiores a 90%.

Ainda, no artigo [26], os autores propuseram um modelo baseado numa RNC para a detecção e contagem de palmeiras. Este método testado contra árvores manualmente identificadas, mostra bons resultados e, uma limitação encontrada estava relacionada com o conjunto de dados. No entanto os autores reportaram resultados superiores a 95%. Outros investigadores propuseram um método com o principal objectivo de generalizar o processo de classificação para diferentes tipos de floresta, utilizando uma rede neuronal em cascata. O estudo reportou resultados promissores, no entanto, os autores pretendem melhorar o seu modelo aplicando uma RNC [15].

As RNC podem ser aplicadas não só à classificação das espécies de árvores, mas também à identificação de algumas características dessas árvores. Isso verifica-se no artigo [12], onde os autores utilizaram as RNC para contar frutos numa

árvore, primeiro identificando regiões potenciais de interesse (a localização dos frutos), depois comparando-a com os dados recolhidos no solo. No fim aplicaram algoritmos de ML com o objectivo de estimar uma contagem final. Os autores propõem a utilização desta metodologia para identificação do fenótipo da planta, contagem de plantas e monitorização de doenças de plantas que apresentem sintomas visuais.

2.5.2 Fundamentos das RNC

Uma RNC, é um algoritmo de ML, que pode aprender analisando grandes quantidades de dados. A sua origem provém de modelos de redes neurais artificiais, que são uma aproximação matemática ao modo de funcionamento de um cérebro humano. O seu processo geral inclui a análise das entradas através de um "neurónio" ou de uma camada de operações matemáticas e a classificação da saída em função do seu valor. Este processo é conseguido através de uma função de activação, a qual serve como um limiar para definir a classe da saída. Para tarefas de reconhecimento de imagem, os valores de entrada são valores de pixels [6, 52].

Esta particularidade é uma vantagem para a análise de imagens, pois pode lidar com a grande quantidade de dados reduzindo a sua dimensão, aliviando assim a carga de processamento computacional. Provaram ser excelentes na extracção de características das imagens, realizando a classificação, segmentação semântica e detecção de objectos [6, 16, 52].

As RNC apresentam uma estrutura básica e os seus principais componentes podem ser classificados de acordo com as suas funções. Os principais componentes serão apresentados nas próximas subsecções e estão ilustrados na figura 2.11. As próximas subsecções basearam-se nos artigos [6, 41].

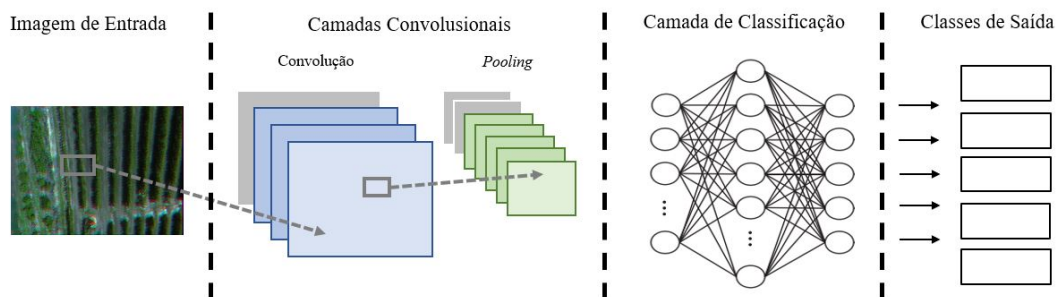


Figura 2.11: Arquitectura de uma RNC, apresentando as principais camadas que a constituem. Imagem baseada em [41].

2.5.2.1 Camada de Convolução

As camadas convolucionais servem como extractores de características e, portanto, elas aprendem as representações das características das imagens de entrada. Os neurónios nas camadas convolucionais estão organizados em mapas de características. O neurónio num mapa de características tem um campo receptivo, que é conectado a uma vizinhança de neurónios na camada anterior por meio de um conjunto de pesos que variam ao longo do treino. As entradas são convolucionadas com os pesos para calcular um novo mapa de recursos e os resultados são enviados por meio de uma função de activação não linear. Todos os neurónios num mapa de características têm pesos iguais; no entanto, mapas de características diferentes dentro da mesma camada convolucional têm pesos diferentes para que as várias características possam ser extraídas em cada local.

Mais formalmente, o mapa de recursos de saída pode ser calculado como

$$Y_k = f(W_k * x) \quad (2.1)$$

Onde a imagem de entrada é denotada por x ; o filtro convolucional relacionado ao k -ésimo mapa de características é denotado por W_k ; o sinal de multiplicação, neste contexto, refere-se ao operador convolucional 2D, que é usado para calcular o produto interno do modelo de filtro em cada localização da imagem de entrada; e representa a função de activação não linear. As funções de activação não linear permitem a extracção de recursos não lineares. Tradicionalmente, as funções sigmóide e tangente hiperbólica foram as mais utilizadas. Recentemente, *rectified linear unit* (ReLU) tornaram-se muito populares. A sua popularidade e sucesso abriram uma área de pesquisa que se concentra no desenvolvimento e aplicações de novas funções de activação para melhorar o desempenho das RNC.

2.5.2.2 Camada de Pooling

O objectivo das camadas de *pooling* é reduzir a resolução espacial dos mapas de características e, assim, alcançar invariância espacial para inserir distorções e traduções. Inicialmente, era prática comum usar camadas de *pooling* médio para propagar a média de todos os valores de entrada, de uma pequena vizinhança de uma imagem para a próxima camada. No entanto, em modelos mais recentes, as camadas de *pooling* máximas que propagam o valor máximo dentro de um campo receptivo para a próxima camada tem ganho destaque.

Formalmente, as camadas de *pooling* máximo seleccionam o maior elemento dentro de cada campo receptivo, de modo que

$$Y_{kij} = \max_{(p,q) \in \mathcal{R}} x_{kpq} \quad (2.2)$$

Onde a saída da operação de *pooling*, associada ao k-ésimo mapa de características, é definido por Y_{kij} , o elemento na localização (p,q) contido na região de *pooling* ij, que incorpora um campo receptivo em torno da posição (i, j), é designado por x_{kpq} .

A Figura 2.12 ilustra a diferença entre o *pooling* máximo e o *pooling* médio. Dada uma imagem de entrada de tamanho 4×4, se um filtro 2×2 e passo de dois for aplicado, o *pooling* máximo produz o valor máximo de cada região 2×2, enquanto o *pooling* médio produz o valor inteiro arredondado médio de cada sub região.

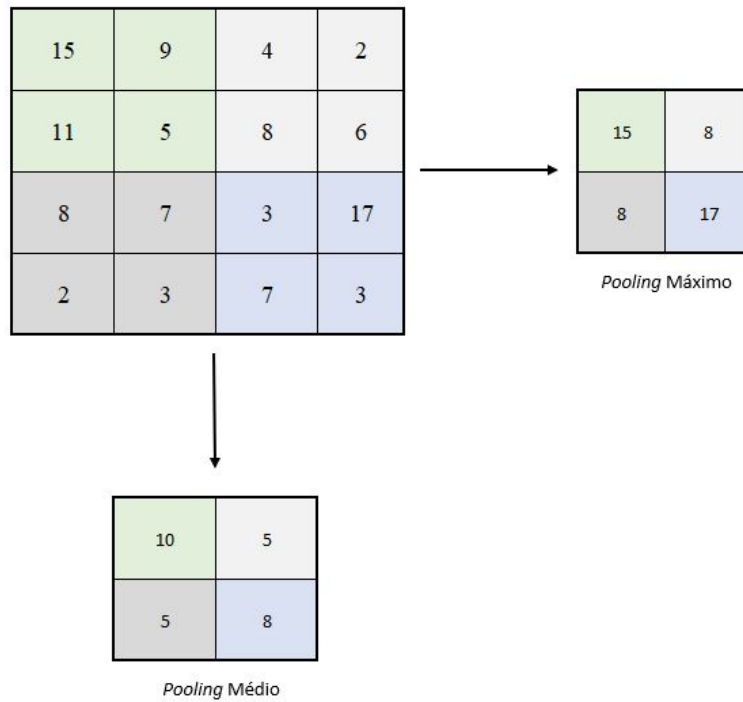


Figura 2.12: *Pooling* máximo e *Pooling* médio. Imagem baseada em [41].

2.5.2.3 Camada de Classificação

Esta camada, também designada por camada totalmente ligada, é a camada que se encarrega de classificar as características extraídas das camadas anteriores e gerar uma saída. Várias camadas convolucionais e camadas de *pooling* são geralmente empilhadas umas sobre as outras com o objectivo de extrair representações das características mais abstractas. As camadas totalmente ligadas que seguem essas camadas interpretam essas representações de características e desempenham a

função de raciocínio de alto nível. Para problemas de classificação, é padrão utilizar o operador *SoftMax* como classificador, no entanto com o desenvolvimento dos algoritmos de classificação, verificou-se que o algoritmo SVM obtinha resultados bastante mais aperfeiçoados. O classificador *SoftMax* é geralmente utilizado para normalizar a probabilidade das anotações, conforme descrito matematicamente na seguinte equação:

$$\text{softmax}(y_i) = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (2.3)$$

Comparar o desempenho de diferentes classificadores de RNCs ainda requer investigação adicional, pois como apresentado anteriormente, a escolha do melhor algoritmo para classificar depende muitas vezes da tipologia dos dados e não do próprio algoritmo.

2.5.3 Arquitecturas de RNCs

Nos últimos anos, várias arquitecturas de RNCs surgiram, e muitos estudos foram feitos, avaliando seu desempenho. Todas elas são baseadas nas camadas estruturais apresentadas anteriormente. A primeira arquitectura de um RNC proposta foi na década de 1990 por LeCun com o nome de *LeNet*, sendo considerada o ponto de partida em termos de arquitecturas de redes aplicadas ao reconhecimento de imagem. Na lista a seguir são apresentadas algumas das arquitecturas mais populares, as descrições da rede basearam-se no artigo [6]:

AlexNet: Proposta em 2012, está preparado para receber imagens com 3 canais, com uma dimensão 256×256 pixels. AlexNet tem 3 camadas de convolução e 2 camadas totalmente conectadas. Foi a primeira arquitectura a utilizar a função de activação *ReLU* em vez de sigmóide, sendo usado após cada camada convolução.

GoogleNet: Proposta em 2014, introduziu novos recursos para a rede, por exemplo, um novo módulo denominado iniciação, que executa diferentes convoluções sobre a mesma camada em paralelo e um método global de média de distribuição que obtém o valor médio de cada mapa de características para alimentar a camada de classificação. Esses novos recursos permitem que a rede tenha mais camadas sem uma grande diminuição no desempenho. Esta rede possui 22 camadas no total.

ResNet: Proposta em 2015, e foi desenvolvida com diferentes números de camadas, começando em 34, 50, 101, 152, chegando até às 1202. No entanto as arquitecturas mais utilizadas eram as de 50 e 152, sendo que até à data de lançamento eram as redes mais profundas. A popular ResNet50 continha 49 camadas

de convolução e 1 camada totalmente conectada no final da rede. Ele introduziu o uso de normalização em lote após cada camada de conversão e um novo módulo denominado bloco residual. Este último consiste em ignorar uma camada ou conjunto de camadas com o valor da camada original, somando-as, esta adição é importante, pois representa uma mudança no valor de entrada.

No entanto, existem muitas outras redes pré-treinadas com diferentes características. A tabela 2.2 apresenta algumas das características de diversas RNCs.

Modelo	Tamanho (MB)	Top-1 Precisão	Top-5 Precisão	Parametros	Profundidade	Dimensão de Entrada
VGG16	528	0,713	0,901	138 357 544	23	224x224
VGG19	549	0,713	0,9	143 667 240	26	224x224
ResNet50	98	0,749	0,921	25 636 712	50	224x224
ResNet101	171	0,764	0,928	44 707 176	101	224x224
InceptionV3	92	0,779	0,937	23 851 784	159	299x299

Tabela 2.2: Características de algumas das redes pré-treinadas mais conhecidas. Tabela baseada em [24].

A figura 2.13 confirma os resultados de precisão apresentados na tabela anterior. Esta figura é o resultado de um estudo de análise as diferentes RNCs pré-treinadas [10].

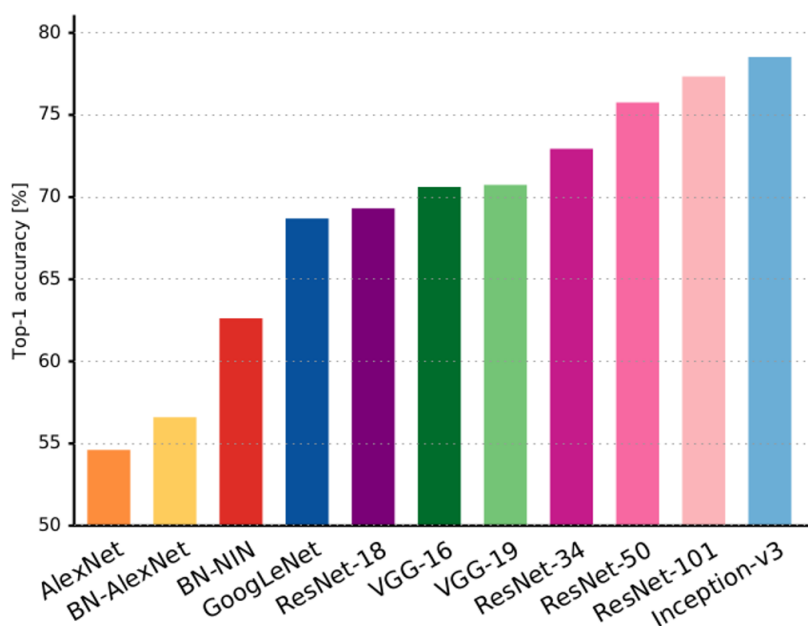


Figura 2.13: Gráfico de comparação de precisão entre diversas RNCs pré-treinadas. Imagem retirada de [10].

2.6 Considerações finais

O recente progresso no desenvolvimento de tecnologias permitiu o aparecimento de novos sensores e novas plataformas, tais como os VANTs. Este tipo de combinações, VANTs equipados com sensores, como câmaras multiespectrais, permitiram desenvolver sistemas autónomos bastante versáteis e flexíveis para diversos cenários de operação.

No desenvolvimento do estado de arte foi possível averiguar as vantagens e desvantagens da utilização dos VANTs quando comparada com os seus competidores, onde foi feita uma breve análise comparativa das diversas características como: autonomia, capacidade de carga, tempo de processamento, resolução, precisão, custo, entre outras. Verificou-se também, que a combinação dos VANTs com sensores de imagem, permite explorar várias áreas de estudo relacionadas com a floresta, tais como: estimativa dos parâmetros dendrómetros, monitorização de incêndios e recuperações pós-incêndio, monitorização da saúde das florestas e, por fim, identificação e classificação de espécies de árvores florestais. Este último, foi abordado com maior detalhe, onde se verificou que através de diversas plataformas, tais como, VANTs, aeronaves e satélites, quando equipadas com sensores de imagens multiespectrais ou hiperespectrais, é possível identificar e classificar espécies de árvores florestais. Durante a investigação realizada, verificou-se que a utilização deste tipo de sensores permitiu obter bons resultados de classificação, tendo um erro muito reduzido. Verificando assim que através do conjunto de dados disponível para classificação era possível desenvolver um modelo que resolvesse o problema proposto nesta dissertação. Foram abordadas diferentes técnicas como métodos de ML e RNCs. As RNCs terão um papel fundamental, pois é através delas que é possível extrair as características que diferenciam as diversas espécies. Foi possível averiguar alguns casos de estudo onde estas redes eram aplicadas para classificação de imagens. Ainda, foi feita uma análise estrutural à arquitetura de uma RNC, onde foi possível perceber o funcionamento e organização de cada camada dessas redes. Os algoritmos de ML serão utilizados para classificar as imagens com base nas características extraídas pela rede. Foram apresentados alguns exemplos de algoritmos de ML, no entanto com base nos estudos, o algoritmo SVM foi o que apresentou mais consistência e desempenho nos resultados de classificação de imagens.

Tendo em conta o conteúdo de toda a secção apresentada anteriormente, será proposta uma solução para a classificação de espécies de árvores nas secções seguintes. Começando por apresentar os métodos e materiais utilizados, seguido da implementação do modelo.

MÉTODOS E MATERIAIS

Neste capítulo, serão apresentados os materiais e métodos usados para resolver o problema proposto. A secção 3.1 descreve os métodos e algoritmos utilizados posteriormente na implementação da solução proposta e a secção 3.2, as especificações dos materiais necessários.

3.1 Métodos

Esta secção descreve os métodos adoptados para resolver o problema de classificação da vegetação. Os métodos são descritos pela ordem de aplicação no modelo proposto, apresentado no capítulo seguinte. Primeiro, serão apresentados os detalhes da RNC e por último, o algoritmo SVM. Ambos os métodos já foram alvos de estudo no capítulo 2, no entanto nesta secção será apresentada com uma abordagem mais formal do funcionamento de ambos os modelos.

3.1.1 Redes Neurais Convolucionais

As redes neuronais convolucionais (RNCs) são uma poderosa ferramenta de ML, formadas pela junção de vários neurónios [16]. Para compreender melhor este conceito de neurónio, é apresentado um neurónio de entrada única na figura 3.1.

De acordo com a figura 3.1 e o artigo [16], o neurónio tem como parâmetro de entrada p . Este é multiplicado pelo peso w , formando wp , um dos termos que é enviado para o somatório. A outra entrada do somatório, 1, é multiplicada por um *bias*, b , e é enviada para o somatório. A saída do somatório n , frequentemente

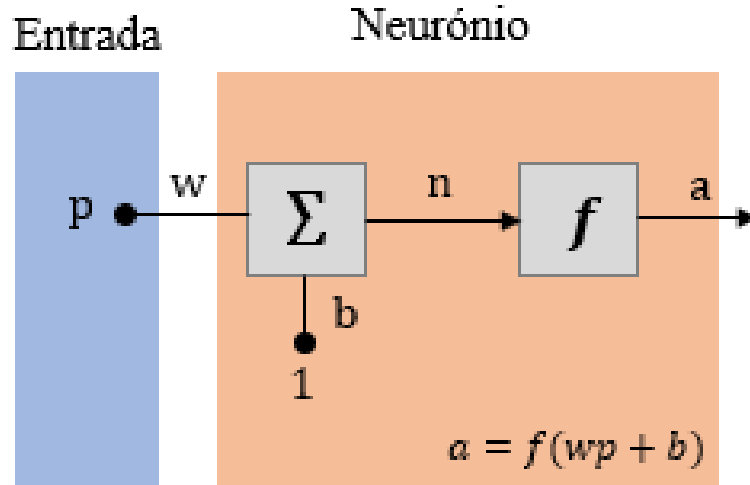


Figura 3.1: Neurónio de entrada única. Imagem baseada em [16].

referida como a entrada da rede, é aplicada a uma função de activação (ou função de transferência), que produz a saída do neurónio, a . A função de saída do neurónio é dada pela equação 3.1.

$$a = f(Wp + b) \quad (3.1)$$

A saída do neurónio depende da função de activação escolhida. Observa-se que W e b são parâmetros ajustáveis do neurónio. Normalmente, a função de transferência é escolhida pelo projectista e os parâmetros w e b serão ajustados para algum objectivo específico. Geralmente, um neurónio tem mais de uma entrada, podendo ter um vector de entradas p ($p_1; p_2; \dots; p_R$) multiplicado por uma matriz de pesos W ($W_1; W_2; \dots; W_R$). Neste caso a função de saída é dada pela equação 3.2.

$$a = f(Wp + b) \quad (3.2)$$

Normalmente, um neurónio, mesmo com diversas entradas, não é suficiente, sendo necessária a combinação de diversos neurónios, trabalhando em paralelo. A esta combinação de neurónios designa-se por camada. Se várias camadas forem combinadas, o resultado será uma rede neuronal. Um exemplo de arquitectura de uma rede neuronal é ilustrada na figura 3.2.

Analisando a figura 3.2, verifica-se que cada camada tem a sua própria matriz de ponderação W , o seu próprio vector de polarização b , um vector de entrada n e um vector de saída a . Como mostrado, existem R entradas, S_1 neurónios na primeira camada, S_2 neurónios na segunda camada e S_n neurónios na N -ésima camada. Camadas diferentes podem ter números diferentes de neurónios e as

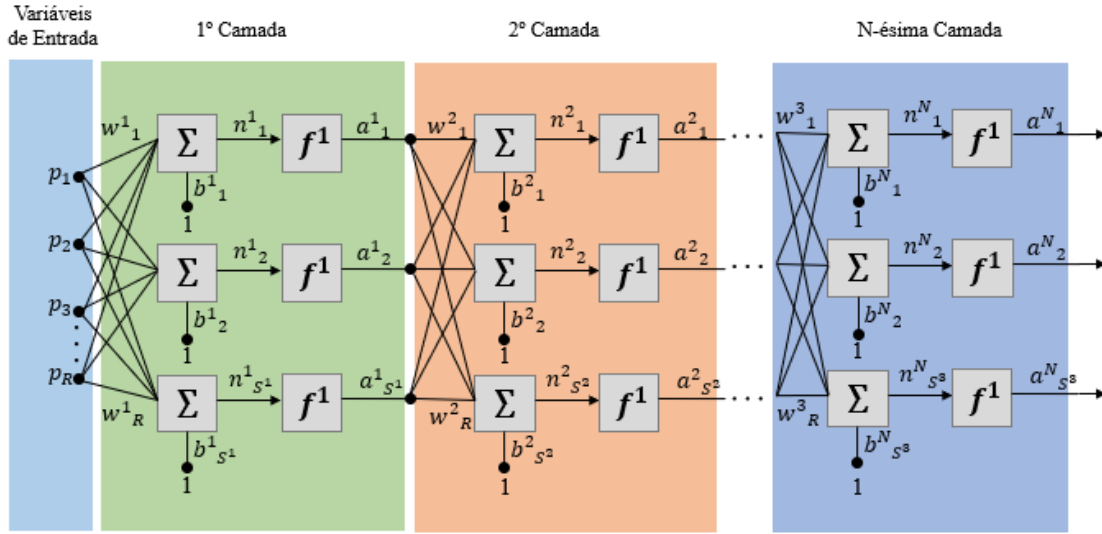


Figura 3.2: Rede Neuronal genérica. Imagem baseada em [16].

saídas das camadas mais baixas são utilizadas como as entradas das camadas mais altas, criando uma rede neuronal. A camada cuja saída é a saída da rede é chamada de camada de saída e as camadas intermédias são chamadas de camadas ocultas.

O objectivo principal da convolução é extrair recursos da imagem de entrada. Convolução é uma operação matemática entre duas funções f e g , resultando numa terceira função que pode ser interpretada como uma função modificada de f (observe que a convolução entre f e g é diferente da convolução entre g e f). No processamento de imagens, onde uma imagem é definida como uma função bidimensional (em imagens em tons de cinza), a convolução é útil na detecção de bordas, suavização de imagens, extracção de características, entre outros [6, 16]. A convolução preserva a relação espacial entre os pixels, extraindo certas características específicas de cada imagem usando pequenos quadrados dos dados de entrada.

Sendo a convolução um processo importante nas RNCs, e no caso em questão a camada responsável pela extracção das características, necessárias será abordado ao detalhe, com o intuito de mostrar como o processo funciona nas imagens de entrada de uma rede.

Sejam f e g duas funções de x , onde x é uma variável contínua. As convolução entre eles são definidas como:

$$f(x) * g(x) = \int_{-\infty}^{+\infty} f(\tau)g(x - \tau)d\tau \quad (3.3)$$

Onde $*$ representa o operador de convolução. Se x for uma variável discreta, a

equação de convolução é dada por:

$$f[x] * g[x] = \sum_{n=-\infty}^{+\infty} f[n]g[x-n] \quad (3.4)$$

Estendendo este conceito para duas variáveis x e y (duas dimensões), são obtidas as seguintes equações:

$$f(x,y) * g(x,y) = \int_{\tau_1=-\infty}^{+\infty} \int_{\tau_2=-\infty}^{+\infty} f(\tau_1,\tau_2)g(x-\tau_1,y-\tau_2)d\tau_1d\tau_2 \quad (3.5)$$

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} f([n_1,n_2])g(x-n_1,y-n_2) \quad (3.6)$$

A convolução de uma imagem com um filtro pode ser interpretada como a soma da multiplicação de cada elemento das imagens, com seu vizinho, pelos elementos da matriz que representa o filtro de convolução, como ilustrado na figura 3.3.

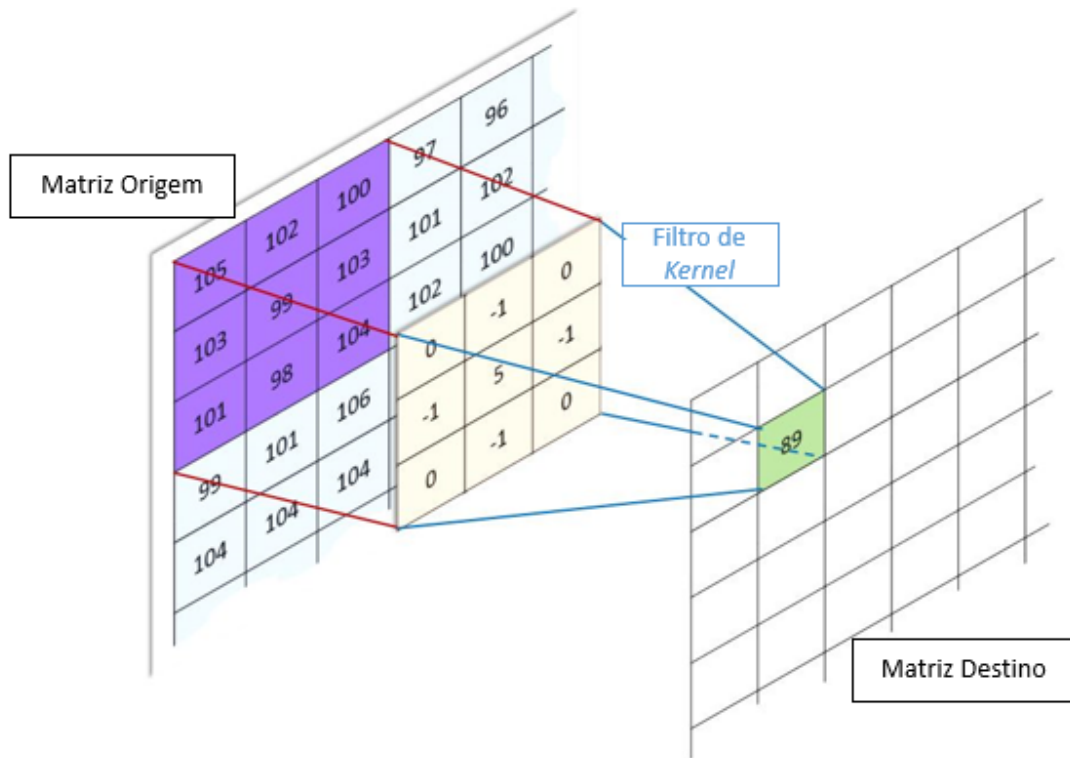


Figura 3.3: Convolução de uma imagem com um filtro 3x3. Imagem baseada em [42].

Em imagens em tons de cinza, onde há apenas um canal, da convolução resulta apenas num mapa de características. Nas imagens multiespectrais, com cinco canais, para realizar a convolução nessas imagens, é efectuada a convolução de cada

canal com um filtro de forma independente na primeira camada de convolução e, a seguir, as saídas são adicionadas. Após a aplicação do filtro, a saída resultante tem o tamanho da imagem original, reduzida por um factor proporcional ao tamanho do filtro. No entanto, existem técnicas para evitar esse problema. Uma das abordagens possíveis é criar pixels ao redor da bordas da imagem de forma que o resultado da convolução tenha o mesmo tamanho da imagem original.

A convolução utiliza pequenas porções da imagem (por exemplo, 8x8 pixels), extraindo características específicas desses pequenos fragmentos da imagem, amostrados aleatoriamente da imagem original. O mapa de características extraído de um fragmento pode ser aplicado em qualquer parte da imagem. Especificamente, as características extraídas podem ser convolvidas com a imagem original, obtendo assim um valor de activação dessa característica em diferentes partes da imagem [6].

3.1.2 *Support Vector Machines (SVM)*

O principal objectivo do algoritmo SVM é encontrar o hiperplano de separação ideal que maximize a margem dos dados de treino. Um exemplo de um hiperplano 2D é mostrado na figura 3.4. Nesse caso, o hiperplano é apenas a linha que divide os dois dados rotulados diferentes [14].

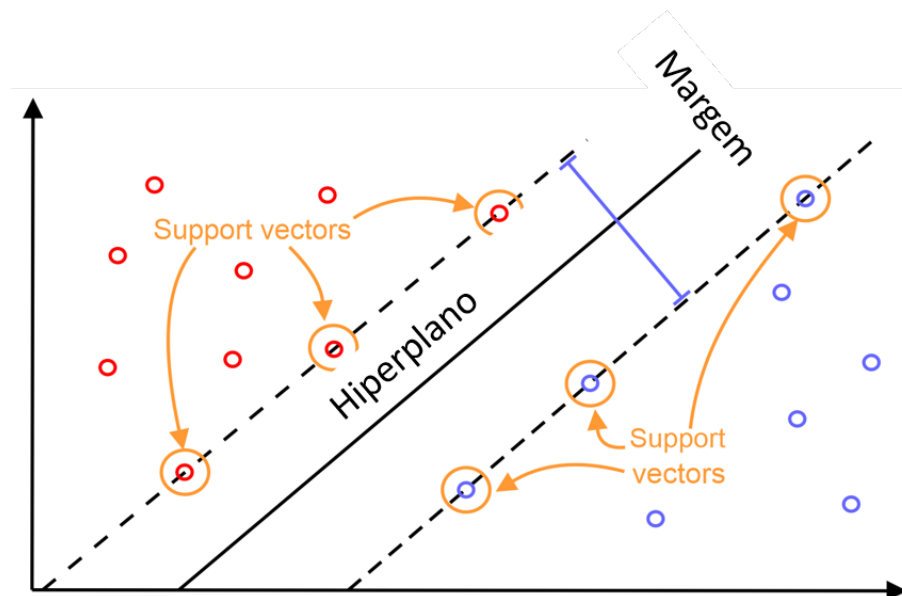


Figura 3.4: Algoritmo SVM. Ilustração do hiperplano que divide os dados de diferentes classes. Imagem retirada de [14].

Dado um determinado hiperplano, a distância entre o hiperplano e o ponto de dados mais próximo pode ser calculada. A margem é obtida duplicando esse

valor da distância calculada. No entanto, esse hiperplano específico pode não ser o ideal. Por definição, o hiperplano óptimo é aquele que maximiza a margem dos dados de treino. Para encontrar o hiperplano ideal, dois planos podem ser seleccionados, separando os dados sem pontos entre eles e a distância entre eles deve ser maximizada. O hiperplano ideal é aquele que se encontra entre esses dois planos. Dessa forma, encontrar a maior margem é o mesmo que encontrar o hiperplano óptimo [14, 49].

O objectivo principal do SVM é produzir um modelo (baseado em dados de treino) que prevê os valores alvo dos dados de teste considerando apenas os atributos dos dados de teste. Seja D um conjunto de treino, composto por n vectores de (características, anotação), $(x_1; y_1), \dots, (x_n; y_n)$. A definição formal de um conjunto de dados inicial é dada pela equação 3.7.

$$D = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in 1, \dots, k_{i=1}^n\} \quad (3.7)$$

Onde k é o número de classes do conjunto de treino. A equação do hiperplano é apresentada na equação 3.8, onde $w.x$ é o produto interno entre os vectores w e x . Sendo $w \in \mathbb{R}^p$, o vector normal para o hiperplano e $\frac{b}{\|w\|}$ é a distância entre o hiperplano e a origem ($b \in \mathbb{R}$).

$$H_0 : f(x) = w.x + b = 0 \quad (3.8)$$

A partir da equação 3.8, é possível obter um número infinito de hiperplanos equivalentes, multiplicando w e b pela mesma constante. Considerando apenas 2 classes alvo, $y_i \in -1, 1$, mais dois hiperplanos podem ser definidos pelas equações 3.9 e 3.10, de modo que H_0 é equidistante de H_1 e H_2 .

$$H_1 : f(x) = wx + b = -1 \quad (3.9)$$

$$H_2 : f(x) = wx + b = 1 \quad (3.10)$$

Para ter certeza de que nenhum ponto está entre H_1 e H_2 , a seguinte restrição deve ser atendida:

$$\left. \begin{array}{l} wx + b \geq 1, \quad \text{se } y_i = 1 \\ wx + b \leq -1, \quad \text{se } y_i = -1 \end{array} \right\} y_i(wx_i) \geq 1 \quad (3.11)$$

$$d = \frac{2}{\|w\|^2} \quad (3.12)$$

Além disso, sabendo que a distância (d) entre H_1 e H_2 (margem) é dada pela equação 3.12, o $\|w\|$ deve ser minimizado para maximizar a margem. Como resultado, um problema de minimização quadrática deve ser resolvido, como se pode ver nas seguintes equações.

$$\min \frac{1}{2} \|w\|^2 \quad (3.13)$$

$$\text{sujeito } a: y_i(wx_i + b) - 1 \geq 0, \forall i = 1, \dots, n \quad (3.14)$$

Como a função a ser minimizada é convexa e a série de pontos em que as restrições são verificadas também é convexa, esse problema tem apenas um mínimo global. Para resolver esse problema, é introduzido o método do multiplicador de Lagrange, resultando na seguinte equação:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_i 1^n a_i (y_i (wx_i + b) - 1) \quad (3.15)$$

A função de Lagrange deve ser minimizada, o que implica maximizar a_i e minimizar w e b .

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n a_i y_i x_i = 0 \quad (3.16)$$

$$\frac{\partial L}{\partial b} = w - \sum_{i=1}^n a_i y_i = 0 \quad (3.17)$$

Substituir as equações 3.16 e 3.17 na equação 3.15 resulta no seguinte problema de otimização dupla:

$$\max \sum_{i=1}^n a_i - \frac{1}{2} \sum_{j=1}^n a_i a_j y_i x_j (x_i x_j) \quad (3.18)$$

Sendo que a equação 3.18 está sujeita as seguintes condições:

$$\begin{cases} a_i \geq 0, \forall i = 1, \dots, n \\ \sum_{i=1}^n a_i y_i = 0 \end{cases} \quad (3.19)$$

No entanto, os padrões nos quais o modelo é treinado podem não ser linearmente separáveis. Neste caso, os dados são mapeados num espaço de características de dimensão superior, usando funções de *Kernel*, definidas por:

$$K(x_i, x_j) = \phi(x_i)^\tau \phi(x_j) \quad (3.20)$$

Kernel é uma função de similaridade, que permite ao SVM operar num espaço de características implícito de alta dimensão, sem calcular as coordenadas dos dados nesse espaço, mas simplesmente calculando o produtos internos entre as imagens de todos os pares de dados no espaço de características, como é ilustrado na figura 3.7.

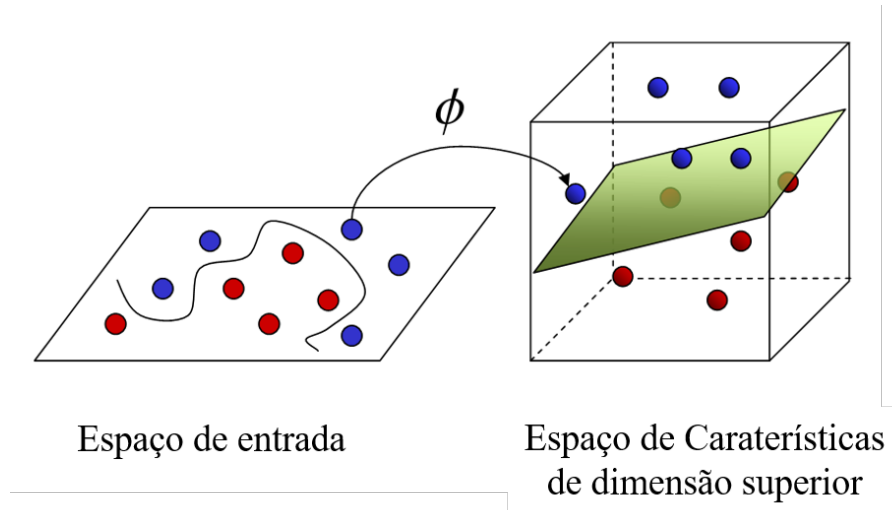


Figura 3.5: Função de *Kernel*: quando os dados não são linearmente separáveis, são mapeados para uma dimensão superior onde podem ser separados linearmente. Imagem baseada em [49].

3.2 Materiais

Para atingir o objectivo desta dissertação os métodos citados são aplicados em imagens aéreas recolhidas por uma câmara montada num VANT. A câmara multi-espectral utilizada é descrita na próxima secção.

3.2.1 VANT

O veículo aéreo utilizado (figura 3.6) é um VANT de 4 motores, com uma dimensão de (716X220X236)mm e aproximadamente 3.8 kg. Este VANT tem um alcance de 7 km e, quando equipado com 2 baterias, permite uma duração de voo superior a 30 minutos dependendo da carga adicional relativa aos sensores. Tem uma capacidade de carga máxima de 2 kg.

O VANT utilizado providencia todo um conjunto de funções de voo autónomas, onde apenas é necessário introduzir alguns parâmetros como a altitude, a velocidade e a área de estudo. Este já vem equipado com diversos sensores como



Figura 3.6: VANT utilizado (DJI Matrice 210 RTK).

GPS e IMU que lhe permite determinar a sua posição e orientação. Além disso foram adicionados alguns componentes ao VANT:

- Sistema Operativo Armbian : esta placa funciona como um mini computador e é responsável pelas operações de software relacionadas com visão computacional e a comunicação via Ethernet ou Wi-fi;
- Câmara RedEdge-M : Câmara multiespectral da MicaSense responsável pela aquisição dos dados.

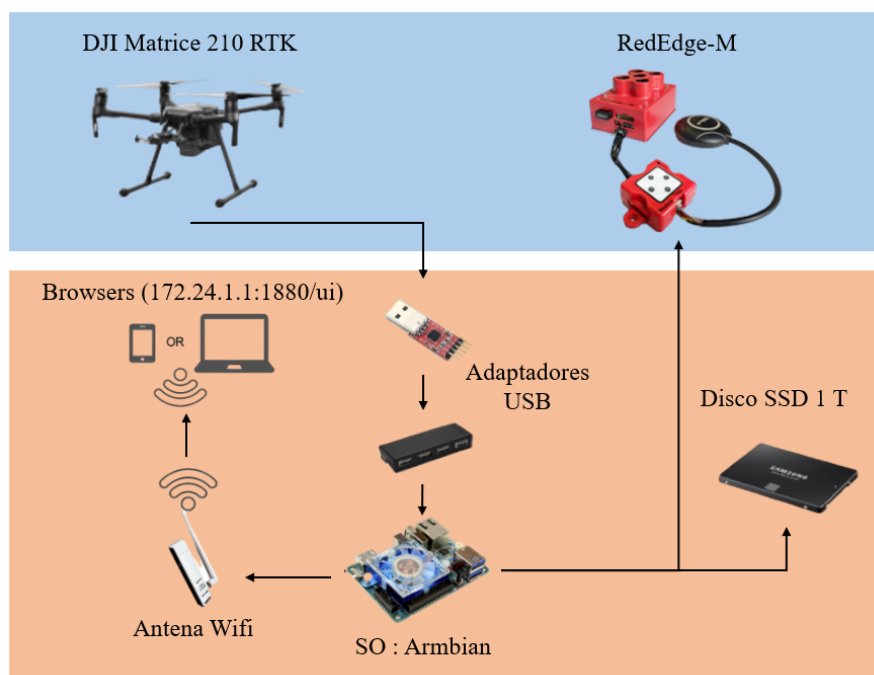


Figura 3.7: Arquitectura do VANT.

3.2.2 Câmara Multiespectral

As imagens utilizadas foram recolhidas pela câmara multiespectral Rededge-M da MicaSense. Este sensor multiespectral possui 5 bandas diferentes, sendo elas as bandas: azul (1), verde (2), vermelho (3) do espectro visível. Bem como as bandas próximo do infravermelho (PIV) (4) e infravermelho (IV) (5). Na tabela 3.1 estão apresentadas algumas características das diferentes bandas do sensor [44].

Número da banda	Nome da banda	centro comprimento de onda	Largura de banda	Alcance
1	azul	475	20	465-485
2	verde	560	20	550-570
3	vermelho	668	10	663-673
4	próximo do infravermelho	840	40	820-860
5	infravermelho	717	10	712-722

Tabela 3.1: Características das bandas capturadas pela câmara RedEdge-M da MicaSense .Tabela baseada em [44].

O conjunto disponibilizado pela MicaSense, além da câmara integra também um sensor de luz, *RedEdge Downwelling Light Sensor* (DLS) e um painel de reflectância calibrado (CRP). O DLS é um sensor de luz incidente de 5 bandas colocado no topo da aeronave ou do VANT, apontando para a frente na direcção de voo. O veículo deve estar o mais paralelo possível ao horizonte (ângulos de rotação e inclinação aproximadamente zero), mantendo uma visão clara do céu durante a missão. Este sensor é responsável por capturar e armazenar a quantidade de informações de luz em cada imagem [32].

Apesar das melhorias contínuas em torno do módulo RedEdge DLS, ele ainda não é um sensor totalmente independente e, por esta razão, é recomendado trabalhar em conjunto com um painel de calibração. Portanto, é importante capturar uma imagem do painel antes e depois de cada missão, para garantir a câmara e as calibrações precisas do DLS para ajustes de sombras, reflexos ou luz [32].

Também deve ser integrado um módulo GPS para receber cada localização da imagem com os dados GPS (latitude, longitude, altitude) e dados de atitude, como os ângulos posicionais da aeronave .

Na figura 3.8 seguinte mostra a integração a integração da câmara proposta da MicaSense conectada aos sensores GPS e DLS.



Figura 3.8: Conjunto MicaSense composto pela câmera RedEdge-M, o DLS e GPS conectados entre si. Imagem retirada de [32].

A câmera RedEdge-M armazena os arquivos num cartão SD numa estrutura de pastas. Uma nova pasta é criada para cada vez que a câmera é ligada. Cada captura é automaticamente armazenada e o formato de imagem pode ser especificado junto com várias outras especificações por meio da página da web da câmera. Cada captura resulta em 5 arquivos DNG (12 bits) ou TIFF (16 bits), diferentes de acordo com a opção de configuração. Um formato de arquivo de 12 bits significa apenas 1,8 MB por imagem, enquanto um formato de 16 bits representa 2,5 MB para cada uma das 5 imagens, embora o segundo formato seja ainda mais recomendado devido à sua maior compatibilidade de software[32].

As imagens do câmara da MicaSense têm resolução de 1280×960 pixels, com proporção de 4:3. De acordo com a figura 3.2, as lentes do sensor possuem um campo de visão horizontal (CVH) de $47,2^\circ$, com uma distância focal de 5,5 mm. A distância da amostra do solo é de 8,2 cm / pixel por banda a 120 m (400 pés) acima do nível do solo[44].

Tipo de Câmera	Multispectral
Nome	MicaSense RedEdge
Comprimento focal (mm)	5.5
Campo de visão horizontal ($^\circ$)	47.2
Bandas espectrais	5
Alcance espectral	465-727
Dimensões	120.9 x 65.8 x 46.0
Tamanho imagem (pixel)	1280 x 960
Profundidade de bits	12 - 16 bit
Formato da imagem	RAW/TIFF
Peso (kg)	0.15

Tabela 3.2: Características da Câmera RedEdge-M da MicaSense. Tabela baseada em [44].

SISTEMA DE CLASSIFICAÇÃO DE VEGETAÇÃO

Este capítulo descreve o fluxo de trabalho adoptado na presente dissertação, ou seja, como os materiais e métodos descritos no capítulo anterior se relacionam e como foram aplicados. A secção 4.1 descreve as imagens aéreas e as características dos dados. A secção 4.2 apresenta o modelo, fazendo uma descrição de todos os processos utilizados. Por último, na secção 4.6, será descrita a visualização dos resultados.

4.1 Apresentação dos Dados

O conjunto de dados, mais especificamente as imagens aéreas, foram recolhidas com a câmara multiespectral da MicaSense, RedEdge-M, apresentada na secção 3.2.2. O conjunto de imagens recolhidas são guardadas num cartão SD, criando novas pastas sempre que a câmara é iniciada. Cada pasta tem a capacidade de suportar até 999 ficheiros, onde cada captura de imagem efectuada origina 5 ficheiros distintos, correspondentes às 5 lentes disponíveis. Essas 5 imagens diferenciam-se entre si através da adição ao nome do ficheiro de um número que varia entre 1 e 5, sendo esses valores correspondentes ao número da banda apresentado na tabela 3.1.

O sensor multiespectral utilizado apresenta funções de calibração das imagens, removendo automaticamente a distorção introduzida pelas lentes. No entanto, o fabricante recomenda a utilização de processos externos de calibração de imagem

pois afirma que o as funções utilizadas pela câmara não apresentam sempre os resultados esperados [32].

Na figura 4.1, é possível verificar que a posição das lentes não se encontra sobreposta, este simples facto faz com que as 5 imagens recolhidas ,simultaneamente, sejam ligeiramente diferentes, pois o ponto onde é feita a captura de imagem é diferente, impossibilitando a combinação dos diferentes canais. A solução consiste em métodos de alinhamento das lentes, ou seja, através de métodos de processamento de imagens criam-se novas imagens mas como se as lentes se encontrassem sobrepostas.

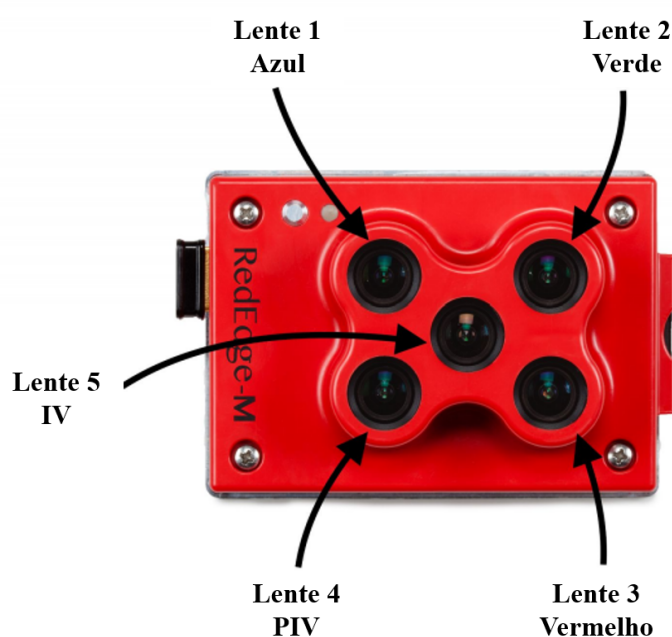


Figura 4.1: Ilustração das diferentes lentes da câmara e a sua correspondente banda espectral. Imagem baseada em [32].

As imagens aéreas foram recolhidas em diversas localizações, a diferentes altitudes, tendo como alvo algumas culturas de arroz, pomares de pereiras, videiras, e tomate. Na tabela 4.1, será possível analisar a quantidade de dados recolhidos, bem como as localizações, divididos pelas 4 classes distintas em estudo.

O conjunto de dados é dividido em 2 subconjuntos. O primeiro subconjunto é designado por conjunto de treino e representa 80% do conjunto de dados. Este é utilizado pela RNC permitindo, através desses dados, desenvolver uma rede que se adapte aos dados e consiga extrair com a melhor eficiência possível as características dos mesmos. Para o treino da rede, este subconjunto ainda é dividido em dois, com uma proporção 80/20, sendo que os 80% representam os dados de treino da rede e os 20% representam os dados de validação da rede. Também

Classe	Nº de Imagens (Após alinhamento)	Nome do Conjunto de Dados
Arroz	86	Arroz(Salvaterra)\2018_09_12_Salvaterra_de_Magos
	542	Arroz(Salvaterra)\2019_07_25_Arroz
total	628	
Pereiras	541	Pera\2019_06_25_Alcobaca
	547	Pera\2019_09_16_Alcobaca
	153	Pera\2020_05_15_Peralta
total	1241	
Tomate	195	Tomate(Roxo e Montalto)\2018_07_24_Roxo
	56	Tomate(Roxo e Montalto)\2019_06_22_Roxo
total	251	
Videiras	270	Vinha(Palmela)\2018_07_27_Palmela
	168	Vinha(Palmela)\2019_08_22_Palmela
total	438	
Total	2558	

Tabela 4.1: Análise quantitativa do conjunto de dados, bem como a sua localização no conjunto de dados.

o algoritmo de classificação SVM utiliza este subconjunto, para adaptar os seus parâmetros às características extraídas pela rede.

O segundo subconjunto, designado por conjunto de teste, e representa os restantes 20% dos dados. Estes dados servem exclusivamente para testar todo o modelo, e calcular os parâmetros de classificação que serão utilizados para representar a validade e eficácia do modelo proposto.

4.2 Modelo Proposto : Visão Global

A abordagem geral do projecto é descrita nesta secção. O objectivo principal era classificar imagens que continham diversas culturas de vegetação, usando a arquitectura de uma RNC pré-treinada. No entanto a rede não suporta imagens multiespectrais, e não está preparada para extrair as características das espécies que seriam utilizadas, nem para classificá-las. Por esse motivo a arquitectura da RNC foi alterada e ajustada aos dados de entrada da rede. As imagens aéreas utilizadas apresentavam uma resolução de 960x1280 pixels, sendo demasiado alta

para os recursos computacionais disponíveis. Visto isto, foi necessário redimensionar as imagens para uma resolução de 480x640 pixels. Após o treino da rede foi desenvolvido um classificador SVM, este utilizava as imagens disponíveis e as características extraídas pela RNC. A RNC também apresenta uma camada de classificação, no entanto, como foi estudado e na secção 2, os algoritmos de ML nomeadamente o SVM apresentavam melhores resultados. Na figura 4.2, está representado todo o modelo de desenvolvimento do classificador proposto.

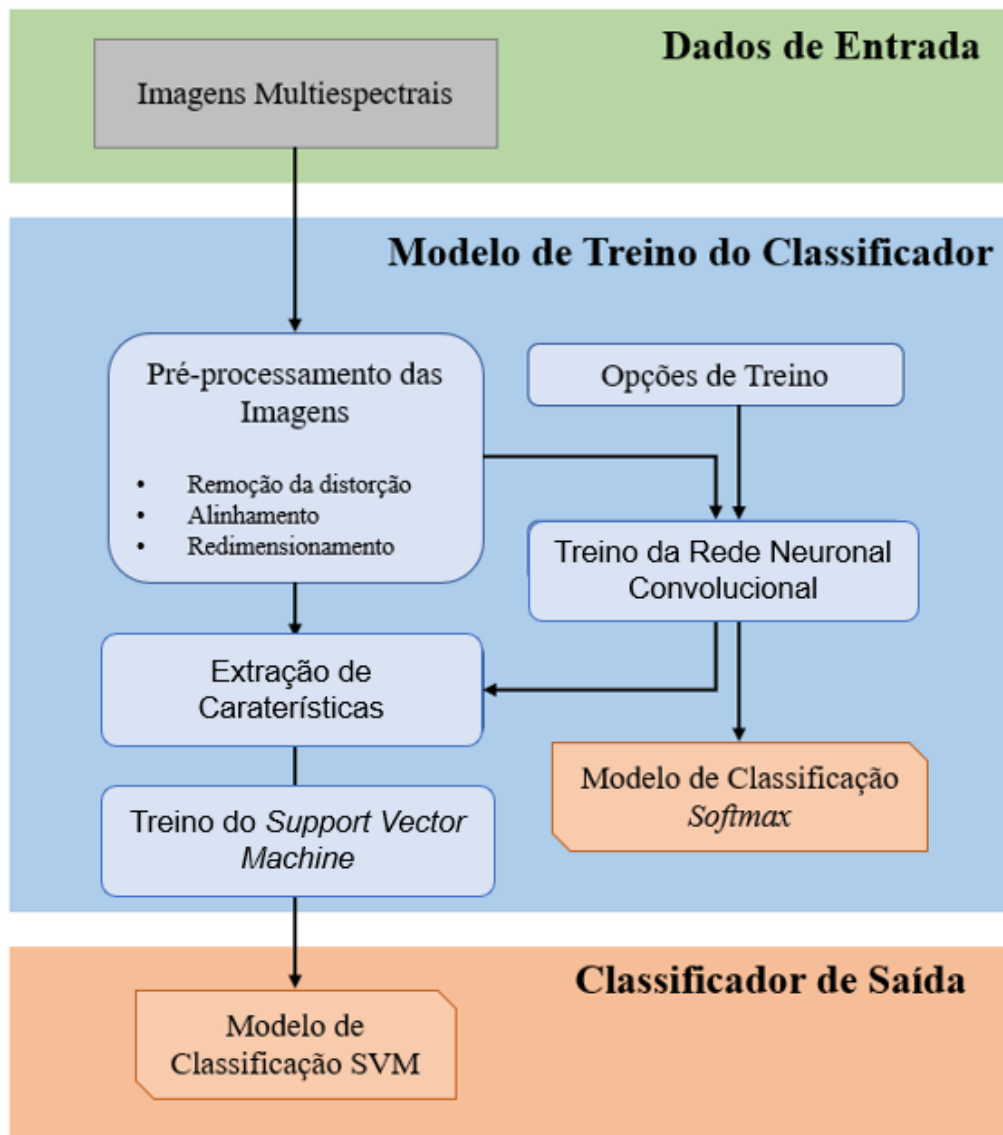


Figura 4.2: Esquemático do modelo de treino.

4.2.1 Ferramentas

Nesta secção, as ferramentas utilizadas são listadas e descritas individualmente:

- *Agisoft Metashape Professional version 1.5.2 (64bits)*, este programa é utilizado para o pré-processamento das imagens aéreas, sendo responsável pela remoção da distorção e alinhamento de todos os canais das imagens.
- *Matlab* foi o ambiente de programação escolhida.
- *Deep Network Designer*, é uma aplicação do *Matlab* que permite alterar as arquiteturas das RNCs, neste caso para alterar algumas camadas da rede escolhida, *Resnet101*.
- *Classification Learner*, é uma aplicação do *Matlab* que permite testar e treinar os algoritmos de ML.
- *Resnet101*, a arquitetura desta RNC foi a escolhida para a extracção das características.
- Todos os processos computacionais realizados utilizaram um computador portátil MSI com um CPU Intel(R)Core(TM) i7-8750H @ 2,20 GHz com 12 núcleos e 16 GB de RAM e equipado com um GPU NVIDIA Quadro P1000 com 4.2GB de memória interna.

Nas secções seguintes, serão apresentados todos processos descritos na figura 4.2, começando por analisar todos os métodos de pré processamento utilizados. Posteriormente será apresentado o modelo de treino da rede RNC e por fim será abordado o classificador SVM.

4.2.2 Pré-processamento de Imagem

O *Matlab* é um programa muito completo, apresentando múltiplas funcionalidades interactivas e oferecendo uma fácil gestão de variáveis. Além disso, possui um pacote muito completo para processamento de imagens, ML e DL, oferecendo múltiplas funções integradas. A maioria das funções que serão necessárias já foram implementadas no *Matlab*, aproveitando a vantagem de estarem muito bem optimizadas, utilizando tanto Unidade Central de Processamento (CPU) como Unidade de Processamento Gráfico (GPU) para processar os dados.

Como mencionado anteriormente, as imagens apresentam alta distorção introduzida pelas lentes, sendo um o processo de remoção de distorção um processo complexo pois cada lente apresenta uma distorção diferente. Antes de qualquer procedimento, todas as imagens do conjunto de dados foram sujeitas á remoção da distorção e alinhamento dos canais. Devido à complexidade desses processos foi utilizado um software comercial especializado em processamento de imagens

aéreas chamado *Agisoft Metashape Professional*. Para remover a distorção, foi necessário calcular os parâmetros intrínsecos da câmara a partir de um conjunto de imagens de um padrão. O padrão, geralmente, utilizado é ilustrado na figura 4.3.



Figura 4.3: Exemplo de um painel de calibração de imagem.

A figura 4.4, pretende demonstrar a importância do alinhamento dos canais. Nesta figura apenas estão representados os 3 canais correspondentes ao espectro visível, no entanto a câmara apresenta 5 canais sendo que os outros 2 não estão representados.

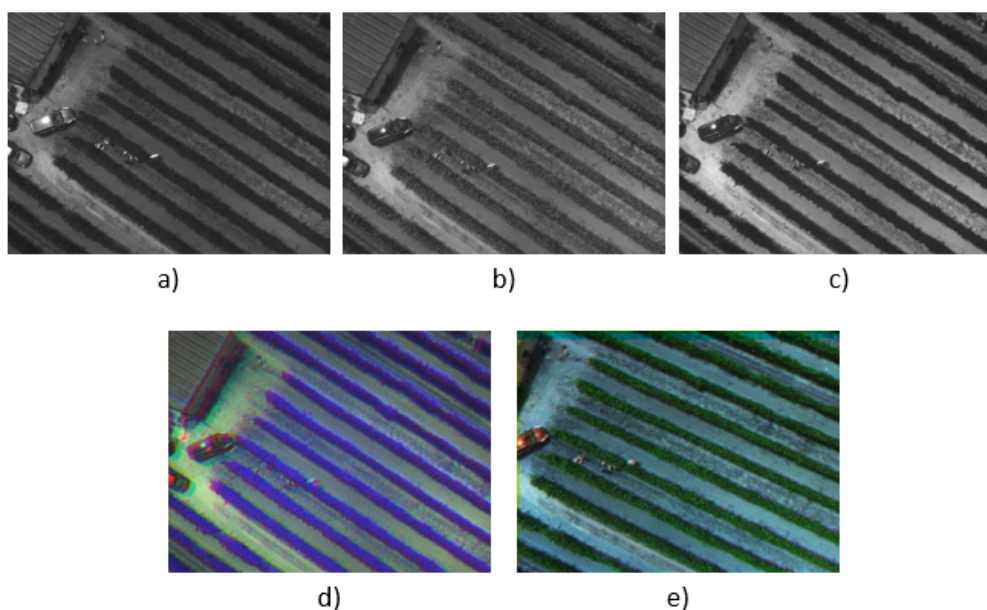


Figura 4.4: a) Canal azul. b) Canal verde. c) Canal vermelho. d) Junção dos 3 canais sem alinhamento. e) Junção dos 3 canais com alinhamento.

Pode verificar-se que, sem alinhamento das imagens, esta encontra-se desfocada que pode ser causada pelo sistema de cores não ser o esperado. Após o alinhamento esses problemas ficam resolvidos, como se pode observar na figura 4.4 imagem e).

A arquitectura da RNC desenvolvida utilizou como variável de entrada imagens multiespectrais com uma resolução de 480x640 pixels. O conjunto de dados original apresentava uma resolução de 960x1280 pixels o que foi necessário redimensionar o tamanho das imagens dos dados. O redimensionamento foi feito utilizando uma função do *Matlab* designada por *augmentedImageDatastore()*. Esta função recebe como argumentos o conjunto de dados num formato de objecto *Imagedatastore* e um vector de 3 posições, sendo eles o comprimento, largura e profundidade da nova dimensão das imagens. O conjunto de dados foi ainda dividido em dois sub conjuntos de dados, como referido anteriormente, utilizando uma função do *Matlab* designada por *splitEachLabel()*, como se pode observar na representação gráfica da figura 4.5. Importante de referir, é o método de divisão dos conjuntos de dados utilizado, uma vez que o conjunto de teste teria que apresentar todas as classes. Para isso foi utilizado o parâmetro *randomized* da função *splitEachLabel()*, que retira um percentagem igual de cada pasta do conjunto de dados.

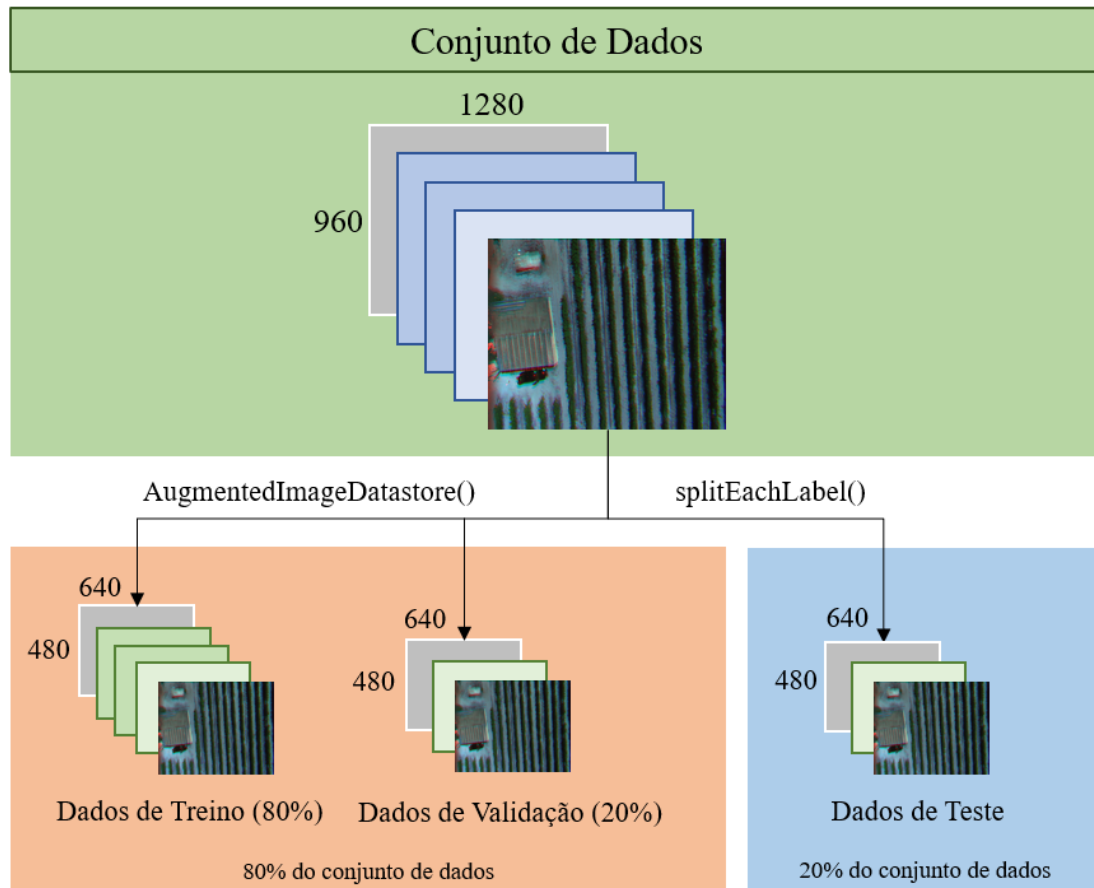


Figura 4.5: Esquema representativo do pré processamento dos dados efectuado no *Matlab*.

Após estarem concluídos estes três processos, o conjunto de dados está pronto para ser utilizado para o treino da RNC e do algoritmo classificador SVM.

4.3 RNC: Arquitectura e Treino da rede

Inicialmente o objectivo delineado para o modelo proposto, passava pela utilização de uma RNC pré-treinada que permitisse, de uma forma fácil e eficiente, extrair características das imagens que facilitassem a sua classificação. No entanto, durante a investigação, não se encontrou nenhuma arquitectura que preenchesse todos os requisitos desejados. A solução encontrada, passou por alterar algumas das camadas de uma rede já existente pois o desenvolvimento de novas redes requer um conhecimento profundo sobre RNCs e tempo, não sendo exequível no tempo de duração desta dissertação.

No seguimento dessa conclusão, foram alteradas e testadas diversas arquitecturas de redes, como por exemplo, *Resnet101*, *Googlenet* e *InceptionV3*. No entanto não se verificou nenhuma alteração nos resultados obtidos pela rede. A escolha da arquitectura da rede responsável pela extracção de características recaiu na rede *Resnet101*.

A rede *Resnet101* apresenta como dimensão das imagens de entrada 224x224 pixels e não estava preparada para receber imagens multiespectrais, apenas suportava 3 canais sendo necessário a alteração de algumas camadas, nomeadamente a camada de entrada. Após a alteração da camada de entrada, a primeira camada de convolução não reconhecia a nova dimensão das imagens de entrada sendo também necessária a sua modificação. Foram também alteradas as camadas de classificação da rede, possibilitando que esta possa classificar as imagens, de forma independente do algoritmo SVM.

Essas alterações foram feitas utilizando aplicação *Deep Network Designer* do *Matlab*. Após as alterações necessárias nas camadas da rede, a aplicação permite exportar a nova rede para um *script*, que permite criar a nova arquitectura da rede sem necessidades de alterar a rede original sempre que necessário.

Contudo ao alterar as camadas da rede esta perde todos os parâmetros pré-treinados pelo autor da arquitectura sendo por isso necessário treinar novamente a rede utilizando os dados de treino e os dados de validação, explicados anteriormente, bem como parâmetros de treino específicos. Na figura 4.6 podemos ver uma representação que ilustra todo o processo de treino.

O processo de treino consiste em duas fase distintas, onde numa primeira fase é feita toda a parte computacional do treino da rede e obtidos os resultados de

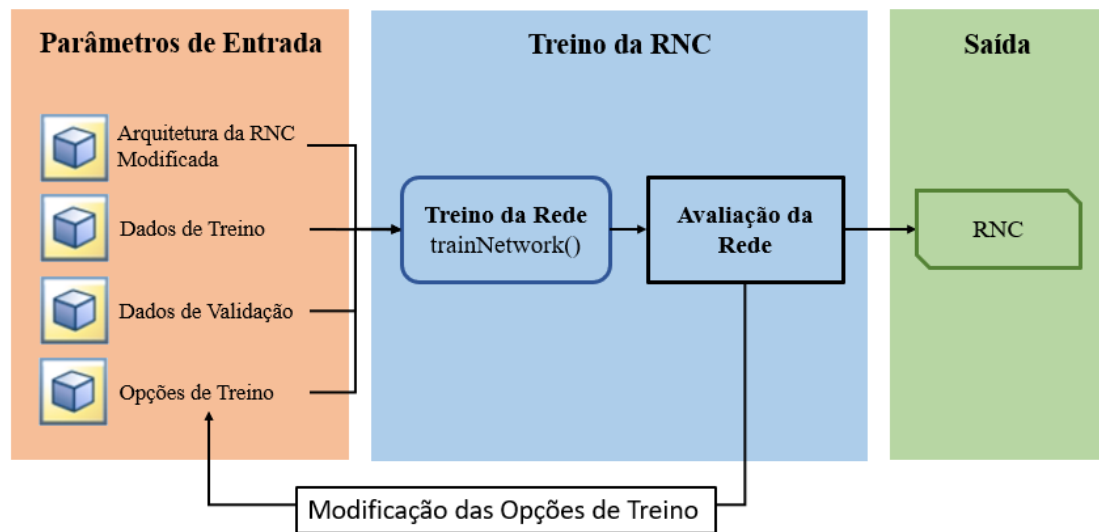


Figura 4.6: Esquema representativo do processo de treino da rede efectuado no *Matlab*

classificação do conjunto de dados de validação. Na segunda fase, é feita uma avaliação dos resultados de classificação obtidos pela RNC na fase anterior. No caso de os resultados serem insuficientes eram alterados os parâmetros de treino e repetia-se novamente o todo o processo. No caso de os resultados serem satisfatórios relativamente aos resultados desejados, o processo de treino fica concluído e a rede está pronta para a extracção das características das classes desejadas. A extracção das características é feita pela camada de convolução mais profunda da RNC, sendo capaz de extrair 4 características diferentes das imagens, características que serão úteis para treinar e testar o algoritmo SVM.

Além de todo processo de treino, o que permite melhorar a eficiência da rede são as opções de treino. Na tabela 4.2, são apresentadas todas as opções utilizadas para o treino da rede.

Recorrendo à documentação [30] e a uma breve análise ao comportamento do treino da rede verificou-se que existiam parâmetros que tinham uma influencia grande nos resultados da rede, sendo esses parâmetros o *MaxEpochs* e o *InitialLearnRate*. Por esse motivo foram variados os valores desses parâmetros com o objectivo de melhorar a eficácia da rede. Foram aplicados 3 valores diferentes, sendo esses valores os que estão apresentados da tabela 4.2 sublinhados a azul. Durante os testes de variação do valor *InitialLearnRate*, o valor do parâmetro *MaxEpochs* era constante e igual a 20. Por outro lado, durante os testes de variação do valor *MaxEpochs*, o valor do parâmetro *InitialLearnRate* era constante e igual a 0.001.

Opções de Treino	
Nome	Valor
<i>SolverName</i>	sgdm
<i>Momentum</i>	0,95
<i>InitialLearnRate</i>	0,01 ; 0,001 ; 0,0001
<i>L2Regularization</i>	0,0005
<i>LearnRateDropPeriod</i>	30
<i>LearnRateDropFactor</i>	0,1
<i>MaxEpochs</i>	10 ; 20 ; 30
<i>MiniBatchSize</i>	7
<i>ValidationData</i>	Dados de Validação
<i>ValidationFrequency</i>	200
<i>ExecutionEnvironment</i>	GPU

Tabela 4.2: Descrição das opções de treino utilizadas.

A utilização de dados de validação permite à rede aplicar o conhecimento adquirido e perceber se a classificação por ela efectuada está correta. Para não ocorrerem problemas de sobreposição, ou seja, as imagens de treino serem iguais as imagens de validação, a rede não utiliza os dados de validação para treino e apenas os aplica, com uma dada frequência, verificando se a sua classificação se adapta aos resultados desejados. Essa frequência é definida pelo parâmetro de treino *ValidationData*, e o valor a ele atribuído corresponde ao número de iterações entre as validações da rede. No entanto, se este valor for demasiado baixo o tempo de treino da rede aumenta.

Como descrito anteriormente, a dimensão das imagens foi reduzida devido à falta de poder computacional disponível. Contudo, mesmo com a nova dimensão das imagens apenas é possível aplicar sete imagens simultaneamente para o treino da rede, utilizando o GPU. A opção de treino que parametriza o número de imagens por iteração é o parâmetro *MiniBatchSize*. Quanto maior for o valor atribuído a este parâmetro, menor será o número de iterações e consequentemente o tempo de treino. Foram testados vários valores para este parâmetro, no entanto, o valor máximo possível foi 7.

4.4 Treino do algoritmo de ML

A utilização de algoritmos de ML em vez da camada de classificação da rede, está relacionado com o facto de os algoritmos não dependerem da rede e poderem ser parametrizados melhorando a eficiência da classificação. Ao contrario da camada de classificação que após o treino da rede não é possível ajustar.

Antes do treino dos algoritmos de ML é necessário desenvolver os mapas de características que permitissem distinguir as diferentes classes. Para a produção dos mapas de características é utilizada a função *activations* fornecida pelo *Matlab*. Esta função utiliza a RNC treinada, como apresentado anteriormente, para extrair os mapas de características do conjunto de dados de treino, através uma camada de convolução definida. Neste caso a camada escolhida foi a ultima camada de convolução que permite extrair 4 características diferentes.

Com o conhecimento dos mapas de características extraídos da rede é possível treinar um algoritmo de ML que permita classificar as imagens em diferentes classes. Porém, as características de cada classe nem sempre são facilmente separáveis de forma linear, sendo por esse motivo, necessário um estudo comparativo dos algoritmos com o objectivo de identificar o algoritmo que obtivesse um melhor desempenho na separação dessas características e consequentemente uma melhor eficiência na classificação das imagens. Para avaliar o desempenho dos vários métodos de classificação, o *Matlab* disponibiliza uma aplicação, designada por *Classification Learner*, que foi utilizada para elaborar um estudo que permitisse obter o algoritmo de ML que melhor se adequava. Para evitar o problema de sobre ajustamento, foi definido na aplicação que 20% dos dados de treino seriam guardados para realizar a validação do modelo do algoritmo. Esta aplicação permite obter diversos gráficos comparativos, sendo os mais conhecidos, a matriz de confusão e a curva ROC.

Após a análise dos algoritmos de ML, o algoritmo seleccionado é treinado com o conjunto de dados completo, ou seja, dados de treino e dados de validação. O treino foi efectuado utiliza a função *fitcecoc* fornecida pelo *Matlab* e tem como variáveis de entrada o mapa de características extraído do conjunto de dados de treino e as anotações das imagens do conjunto de dados de treino. Posteriormente, o modelo desenvolvido é guardado para ser utilizado no processo de teste. Na figura 4.7 podemos ver um esquemático representativo que ilustra o processo de treino do algoritmo.

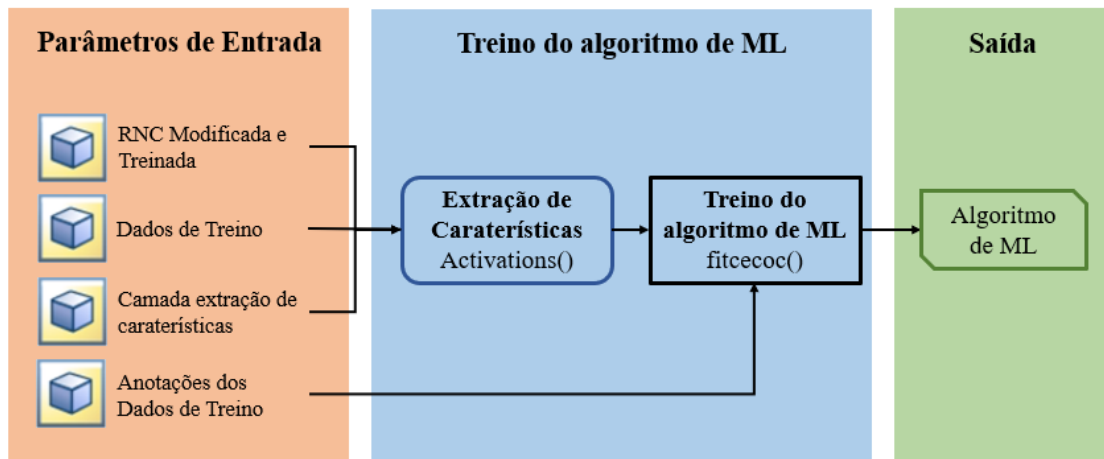


Figura 4.7: Esquemático representativo do processo de treino do algoritmo de ML efectuado no *Matlab*

4.5 Teste do algoritmo de ML

O teste do algoritmo de ML utiliza um conjunto diferente de imagens aéreas, este conjunto foi descrito anteriormente como conjunto de teste. Idealmente, esse conjunto de imagens deve conter várias imagens com as diferentes espécies de vegetação. A fase inicial do processo é idêntica ao processo de treino, pois é necessário desenvolver o mapa de características do conjunto de imagens de teste. No entanto, este processo começa por carregar as imagens de teste, posteriormente é utilizada a função *activations* para extrair as características dessas imagens, através da RNC. A camada de extracção das características é a mesma que foi utilizada no processo de treino, ou seja, é seleccionada a ultima camada de convolução. Após a extracção do mapa de características, é utilizado o algoritmo treinado anteriormente para separar essas características por classes, ou seja, para classificar as imagens por classes. Para isso foi utilizada a função *predict* fornecida pelo *Matlab*, que recebe o mapa de características extraído das imagens de teste e o algoritmo de classificação desenvolvido. Por fim são guardados os dados da classificação das imagens para, posteriormente, serem analisado. Na figura 4.8 está representado um esquemático que ilustra todo o processo de treino, bem como as suas variáveis de entrada e de saída.

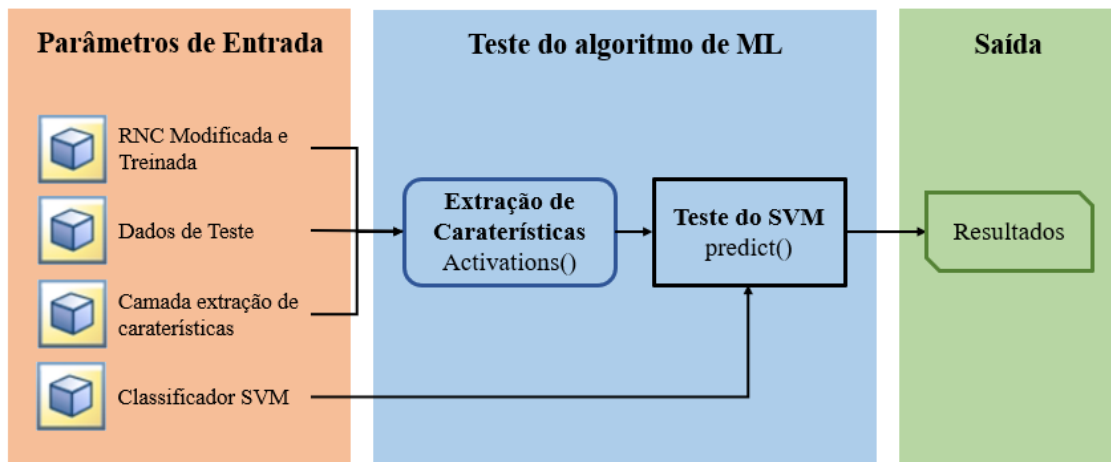


Figura 4.8: Esquemático representativo do processo de teste do algoritmo de ML efectuado no *Matlab*

4.6 Visualização dos resultados

A visualização e apresentação dos resultados é um processo importante pois permite utilizar o classificador para a resolução de outros problemas ou até a permitir ser utilizado como um sub sistema. Para isso, foi desenvolvido um código *Matlab* que permite que seja trabalhados e exportados todos os dados relativos ao classificador, mais propriamente a classificação efectuada, medidas de desempenho do classificador, e as imagens classificadas por classes. Inicialmente, é criada uma nova pasta com o nome de Resultados onde serão guardadas todas essas informações. Dentro dessa pasta serão guardadas as imagens classificadas divididas por pastas, sendo que cada pasta tem o nome da classe a que esta pertence. Numa outra pasta, é guardada uma tabela de classificação, onde apresenta de forma analítica a classificação atribuída a cada imagem e uma tabela onde estão disponíveis os medidores de desempenho utilizados de cada classe. Ainda, são exportados a matriz de confusão e o gráfico *ROC curve* referentes ao desempenho do classificador.

RESULTADOS EXPERIMENTAIS

Este capítulo está dividido em duas partes, numa primeira abordagem é descrito o plano de teste aplicado ao modelo desenvolvido, E na segunda parte, são descritas a aplicação desse plano de testes e os resultados obtido.

5.1 Plano de Teste

Esta secção descreve como foram planeados os testes ao modelo e quais as métricas de desempenho utilizadas. Primeiramente, na secção 5.1.1, são detalhadas as medidas estatísticas adoptadas e suas definições. Na secção 5.1.2, são mostrados os dados seleccionados para testar o sistema. Por último, são descritos os planos de teste da RNC e do algoritmo SVM.

5.1.1 Sistema de Medição de Desempenho

A avaliação de desempenho dos algoritmos de classificação de ML é baseada em medidas estatísticas. A principal medida estatística é designada por exactidão (*accuracy*) e é definida pelo número de previsões corretas dividido pelo total de previsões efectuadas. Neste ponto, é relevante definir alguma nomenclatura e algumas definições das medidas estatísticas.

- Verdadeiro Positivo (Vp): número de amostras que pertencem a uma determinada classe e são classificadas como pertencendo a essa classe.
- Verdadeiro Negativo (Vn): número de amostras que não pertencem a uma determinada classe e são classificadas como não pertencentes a essa classe.

- Falso Positivo (Fp): número de amostras que não pertencem a uma determinada classe, mas são classificadas como sendo dessa classe.
- Falso Negativo (Fn): número de amostras que pertencem a uma determinada classe, mas são classificadas como não pertencentes a essa classe.

Essas medições por si só não dizem muito sobre o desempenho de classificação, mas podem ser combinadas matematicamente com outras medições, como exactidão, precisão, Taxa de Verdadeiro Positivo (TVP), Taxa de Falso Negativo (TFN) ou Taxa de Falso Positivo (TFP), calculada da seguinte forma:

$$exactidao = \frac{Vp + Fn}{Vp + Vn + Fp + Fn} \quad (5.1)$$

$$precisao = \frac{Vp}{Vp + Fp} \quad (5.2)$$

$$recall = TVP = \frac{Vp}{Vp + Fn} \quad (5.3)$$

$$TFN = \frac{Fn}{Fn + Tp} = 1 - TVP \quad (5.4)$$

$$TFP = \frac{Fp}{Fp + Tn} \quad (5.5)$$

Também é comum representar o desempenho graficamente como uma matriz de confusão ou como uma curva de características operacionais do receptor, mais conhecida por *ROC curve*. Numa matriz de confusão, cada linha da matriz representa as instâncias numa classe prevista, enquanto cada coluna representa as instâncias numa classe real. A *ROC curve* é criada através taxa de verdadeiro positivo (TVP) em função da taxa de verdadeiro negativo (TVN).

5.1.2 Conjunto de dados de teste

Para avaliar o desempenho do modelo proposto, foi criado um conjunto de dados de imagens aéreas específico para testar o algoritmo. Como apresentado anteriormente, este conjunto de dados corresponde a 20% do conjunto total e foi dividido de forma aleatória, mantendo as proporções do número de imagens por classe. Este conjunto de dados contém 516 imagens aéreas, divididas por as 4 classes de classificação, arroz, pereiras, tomate e videiras. A tabela 5.1 apresenta o conjunto de dados dividido por classes.

Classe	Nº de Imagens (Após alinhamento)
Arroz	126
Pereiras	248
Tomate	54
Videiras	88
Total	516

Tabela 5.1: Descrição do conjunto de dados de teste.

5.1.3 Teste da RNC

Como apresentado no capítulo anterior, foi necessário alterar a arquitectura da rede, nomeadamente a camada de entrada, e por esse motivo foi necessário treinar uma nova RNC, que permitisse extrair as características que diferenciavam as espécies. Para isso foi alterada também a camada de classificação da rede de modo a ser possível a própria rede classificar as imagens. No entanto esta classificação apenas serviu para indicar a condição da rede e se esta estava preparada para a extracção das características desejadas, pois não era possível parametrizar a camada de classificação nem alterar o algoritmo classificador por ela utilizado. Desse modo, essa classificação não respondia aos critérios desejados uma vez que se pretendia analisar qual seria o melhor algoritmo de classificação.

O sucesso do treino da rede está relacionado com o parâmetros de treino. Os parâmetros de treino com maior influencia no treino da rede foram o número de iterações e a taxa de aprendizagem inicial. Através destes parâmetros ajustou-se a rede de modo a que esta apresentasse uma maior eficácia na extracção do mapa de características utilizado posteriormente pelo algoritmo de ML.

5.1.4 Teste do algoritmo SVM

Para testar o desempenho do algoritmo SVM, foram realizados dois testes. O primeiro método utilizado foi a validação de *Hold-out* dos dados de treino e o segundo processo realizado manualmente usando as imagens do conjunto de dados de teste.

Conforme mencionado na secção 4.5, o *Matlab* possui uma aplicação capaz de testar e comparar a precisão de diferente método de classificação dos dados de

treino. A aplicação utiliza os dados de treino e separa-os em duas partes: uma para o treino do algoritmo e outra para validação. A separação pode ser alcançada usando validação cruzada (geralmente aplicada em pequenos conjuntos de dados) ou validação *Hold-out* (normalmente usada para grandes conjuntos de dados). Como o conjunto de dados de treino apresenta 1652 amostras, foi escolhido o método *Hold-out*, guardando 20% dos dados de treino para validação. O objectivo desta etapa é avaliar a precisão dos métodos de ML, sem usar dados de teste, com o intuito de escolher o que se melhor adequava ao problema. Após a selecção do método de classificação, todo o conjunto de dados de treino é utilizado para fazer o treino real do método de classificação escolhido.

Para o teste do algoritmo SVM foi utilizado o conjunto de dados de teste guardado apenas para o teste final do classificador. Sendo depois aplicado ao resultado da classificação uma análise estatística do desempenho do classificador, utilizando as medidas estatísticas apresentadas anteriormente.

Neste momento, é relevante definir exactamente o que é considerado Vp, Vn, Fp, Fn.

- Verdadeiro Positivo (Vp): classificação da imagem é igual à classe pretendida e a anotação da imagem é igual à classe pretendida;
- Verdadeiro Negativo (Vn): classificação da imagem não é igual à classe pretendida e a anotação da imagem não é igual à classe pretendida;
- Falso Positivo (Fp): classificação da imagem não é igual à classe pretendida mas a anotação da imagem é igual à classe pretendida;
- Falso Negativo (Fn): classificação da imagem é igual à classe pretendida mas a anotação da imagem não é igual à classe pretendida;

5.2 Avaliação e Discussão dos Resultados

Esta secção descreve os resultados alcançados em relação à classificação da vegetação usando os testes descritos na secção anterior. Este capítulo está dividido em três secções principais: a primeira avalia o desempenho da RNC e a segunda mostra o desempenho do SVM, utilizando as medidas estatísticas mencionadas e, utilizando os dados de validação. Por fim, é mostrado o desempenho de todo o modelo utilizando o conjunto de dados de teste.

5.2.1 Conjunto de dados

Antes de qualquer classificação, conforme mencionado anteriormente neste documento, é necessário alinhar as imagens capturadas pelas diferentes lentes da câmara multiespectral. A figura 5.1 ilustra os piores cenários relacionados com os alinhamentos e a remoção da distorção das lentes. Nas imagens verifica-se um alinhamento impreciso não permitindo que a imagem seja correctamente entendida e posteriormente processada. Esta imprecisão contribui para uma degradação do desempenho do modelo nas próximas etapas.

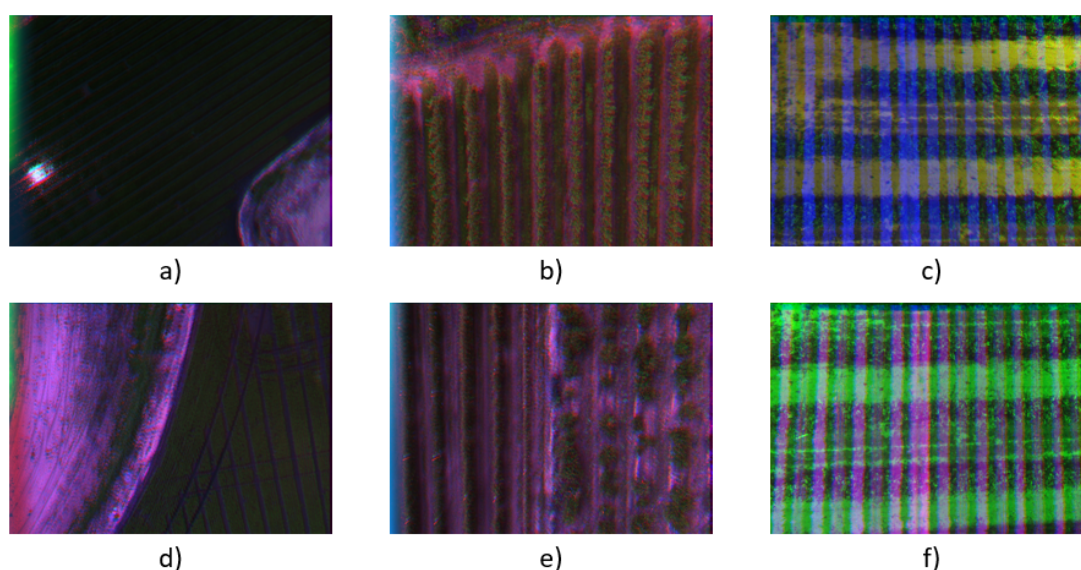


Figura 5.1: Alguns exemplos de falhas no alinhamento e na remoção da distorção das imagens.

A altitude de recolha dos dados está compreendida entre os 40 e 120 metros. Verificou-se que em alguns casos o valor da altitude seria demasiado alto, tornando os principais elementos da imagem, neste caso as espécies de vegetação, demasiado pequenos. Este problema influencia as etapas seguintes do modelo, numa primeira fase na extracção das características, uma vez que, o que permitem a classificação das imagens são as características diferentes de cada espécie. Numa segunda fase, na classificação das imagens, uma vez que o classificador utiliza essas características para determinar a classe de vegetação presente na imagem.

Os dados que não se apresentavam nas melhores condições, mesmo após o pré-processamento, foram mantidos no conjunto de dados pois o objectivo era criar um algoritmo de classificação totalmente autónomo. Por outro lado, uma vez que a extracção de características, executada pela RNC, é feita a cada canal

individualmente não teria um peso demasiado negativo no desempenho do classificador.

5.2.2 Rede Neuronal Convolutional

A verificação do sucesso do treino da RNC foi feita através da classificação do conjunto de dados de validação, apresentado no capítulo 5.2.1. Para a visualização dos resultados foi alterada a camada de classificação da rede permitindo aferir se a rede estava preparada para extrair as características necessárias para o treino do algoritmo de classificação, como descrito na secção 4.3.

Como descrito no plano de teste na secção 4.3, foram ajustados dois parâmetros com o intuito de melhorar a eficiência da rede. O primeiro parâmetro foi o número de épocas (*MaxEpochs*), este parâmetro corresponde ao número de vezes que o conjunto de dados é utilizado no treino da rede. A rede foi testada com 3 valores diferentes desse parâmetro, 10 épocas, 20 épocas e 30 épocas. Na figura 5.2 podemos analisar a resposta da rede a alteração do número de épocas.

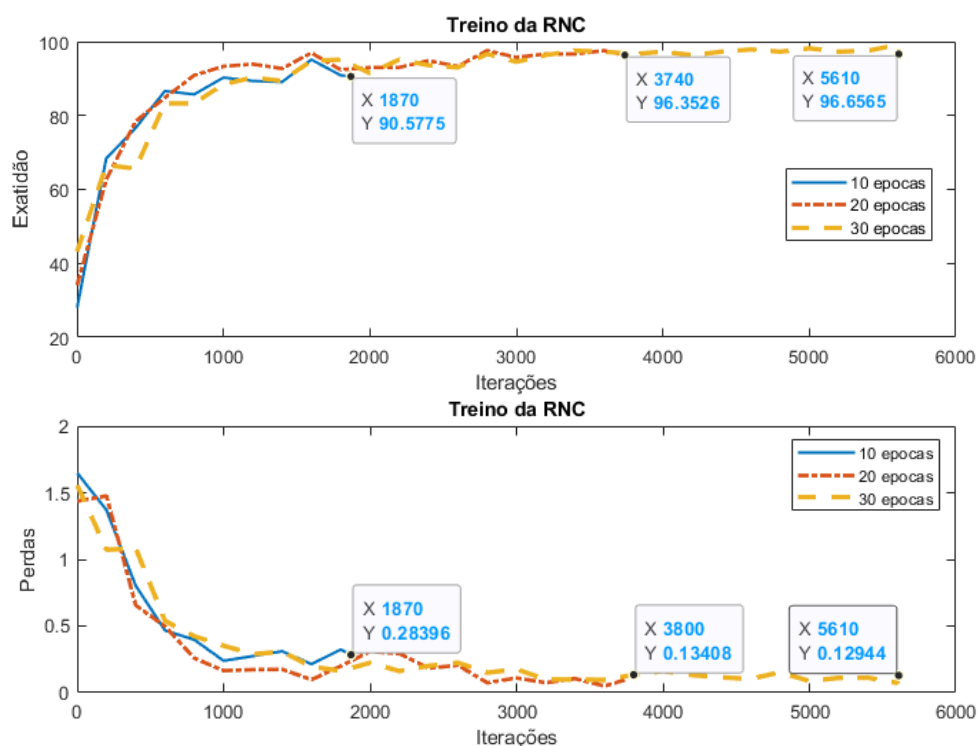


Figura 5.2: Gráficos de comparação do treino da rede com a alteração do parâmetro das opções de treino *MaxEpochs*.

Verifica-se que com o aumento do número de épocas o exactidão da rede fica maior, no entanto quanto maior foi o número de épocas maior é o tempo necessário para treinar a rede, como está descrito na tabela 5.2 .

5.2. AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS

Nº Epocas	Nº Iterações	Exatidão	Perda	Tempo de Treino
10	1870	90,58%	0,283	02:38:54
20	3740	96,35%	0,095	05:28:01
30	5610	96,66%	0,129	07:23:01

Tabela 5.2: Tabela comparativa dos três teste efectuados com os diferentes valores do parâmetro das opções de treino *MaxEpochs*.

O segundo parâmetro de treino alvo de estudo foi a taxa de aprendizagem inicial (*InitialLearnRate*), este parâmetro regula a velocidade de variação dos pesos da rede. Se os valores deste parâmetro forem baixos a duração do treino da rede é muito elevado. Por outro lado, se o valor desse parâmetro for alto a duração do treino é menor, pois a rede aprende mais rápido, no entanto, se for excessivamente alto esta pode divergir. A rede foi testada com 3 valores diferentes desse parâmetro, 0.01, 0.001 e 0.0001. Na figura 5.3 podemos analisar a resposta da rede à alteração do valor da taxa de aprendizagem inicial.

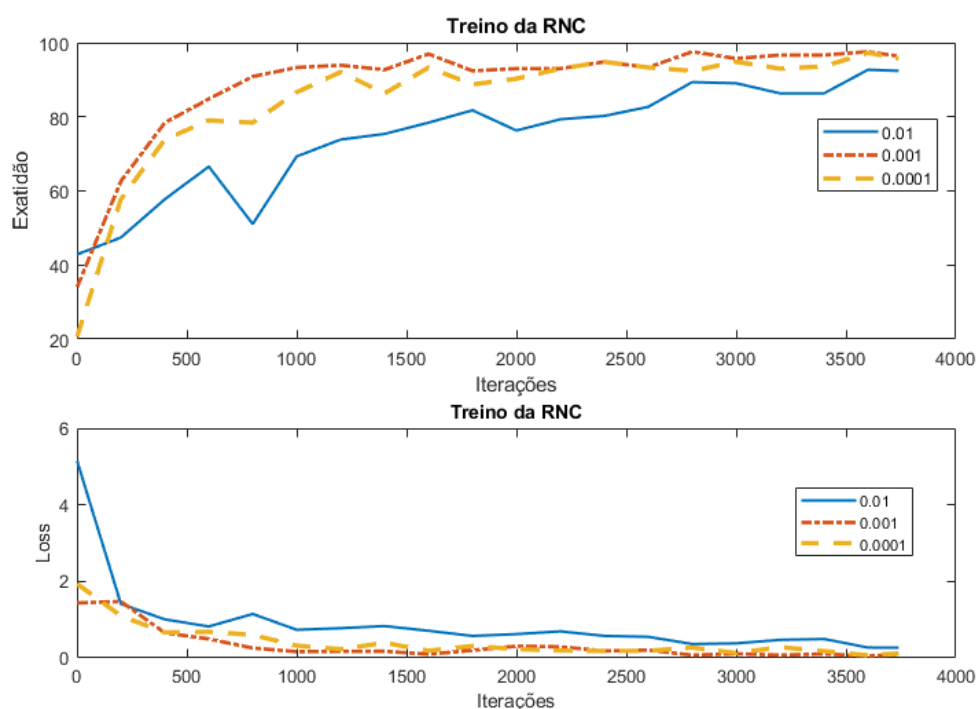


Figura 5.3: Gráficos de comparação do treino da rede com a alteração do parâmetro das opções de treino *InitialLearnRate*.

Verifica-se que com o valor ideal para a taxa de aprendizagem é 0.001, pois como se pode observar na figura 5.3 é o que apresenta um melhor desenvolvimento da rede ao longo das 20 épocas. No entanto denota-se que o resultado final dos três valores em termo de exactidão é idêntico. Na tabela 5.3, é verifica-se

algumas características comparativas entre os valores atribuídos a este parâmetro, mais concretamente o tempo de treino uma vez que a exactidão final é idêntica.

Taxa de Aprendizagem Inicial	Exatidão	Perdas	Tempo de Treino
0,01	92,40%	0,2678	04:58:53
0,001	96,35%	0,0957	05:28:53
0,0001	95,74%	0,1187	05:16:14

Tabela 5.3: Tabela comparativa dos três teste efectuados com os diferentes valores do parâmetro das opções de treino *InitialLearnRate*.

Após a análise destes dois parâmetros foram escolhidas como opção de treinos os parâmetros sublinhados a azul nas tabelas 5.2 e 5.3. A lista de parâmetros escolhidos encontra-se na figura 4.6.

5.2.3 Algoritmo SVM

Iniciando o plano de testes descrito anteriormente, foi utilizada a aplicação *Classification Learner* do *Matlab* para realizar testes em vários algoritmos de ML com vários tipos de kernels e o polinomial cúbico apresentou os melhores resultados com uma precisão de 98,7%. Na figura 5.4 mostra as características extraídas da RNC onde se pode verificar a dificuldade em dividir os dados em classes.

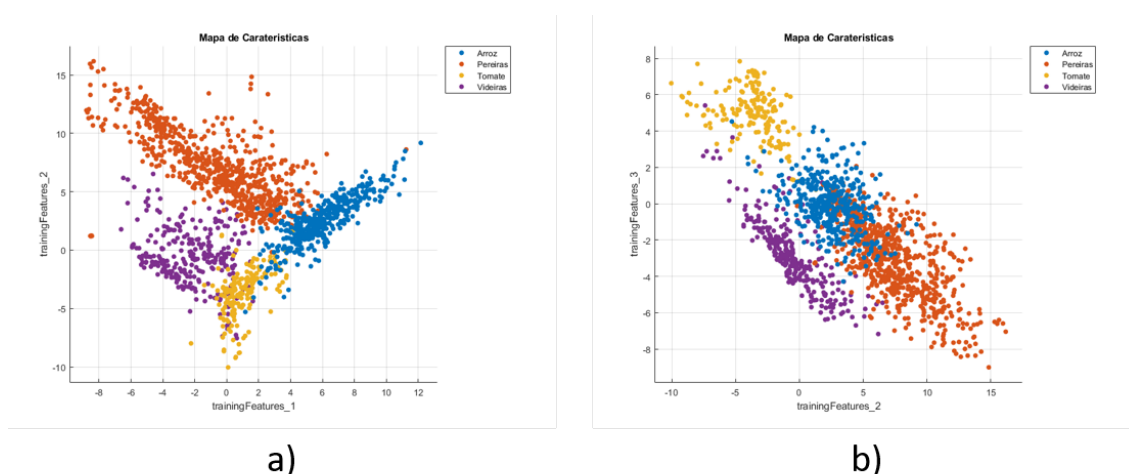


Figura 5.4: Exemplos dos mapas de características extraídos da RNC.

No entanto, os resultados foram superiores aos resultados esperados, como é possível observar na tabela 5.4, onde são apresentados os melhores classificadores

5.2. AVALIAÇÃO E DISCUSSÃO DOS RESULTADOS

obtidos. Como se previa após a pesquisa elaborada no capítulo 2 o algoritmo SVM apresentou os melhores resultados.

Algoritmo	Exatidão
Quadratic SVM	95,80%
Cubic SVM	98,20%
Fine Gaussian SVM	96,70%
Medium Gaussian SVM	96,70%

Tabela 5.4: Comparação entre os melhores algoritmos de ML testados pela aplicação *Classification Learner*.

A figura 5.5 mostra os resultados gráficos do processo de validação do algoritmo *Cubic SVM*. No lado esquerdo, a matriz de confusão mostra que, de 330 amostras (20% de 1652), apenas 6 amostras foram classificadas incorrectamente. No lado direito, a figura 5.5b mostra a curva ROC resultante de um classificador ideal.

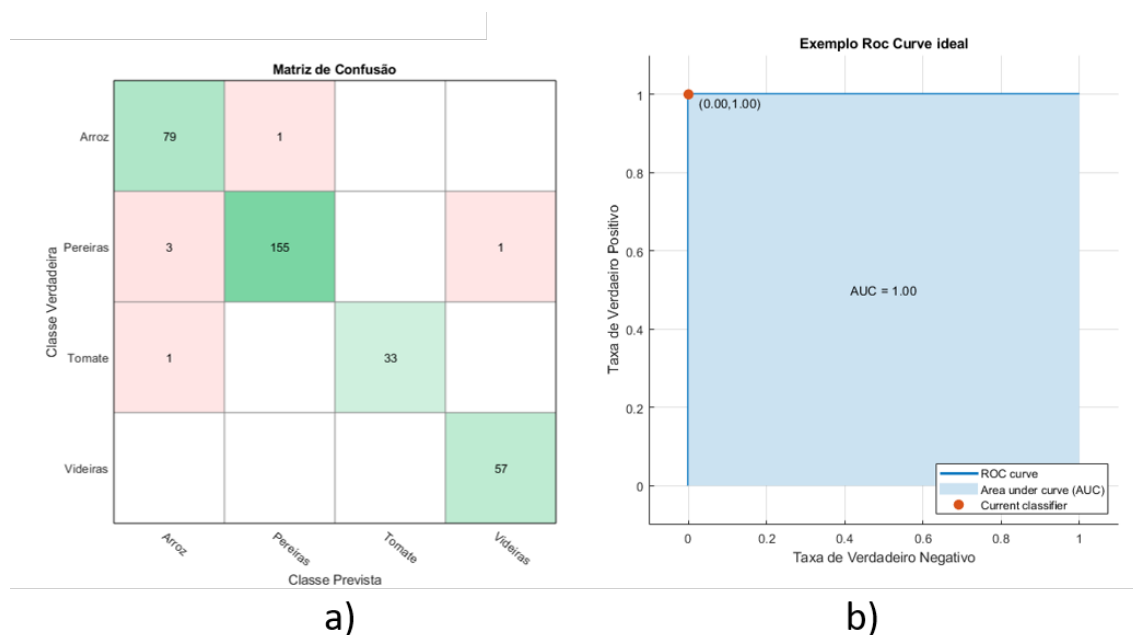


Figura 5.5: Representação gráfica dos resultados do processo de validação. a) Matriz Confusão. b) Curva ROC ideal.

Na matriz de confusão (figura 5.5a), cada linha representa a classe real e cada coluna representa a estimativa do algoritmo SVM. Em cada quadrado, está representado o número de amostras que respeita as condições impostas pelos dois eixos. A diagonal verde representa as boas classificações, ou seja, os verdadeiros

positivos. Os vermelhos representam as classificações erradas. Por exemplo, na primeira classe de imagens prevista, que contém apenas arroz, foram classificadas 79 imagens correctamente como arroz, 3 foram classificadas como pereiras e 1 como tomate.

A tabela 5.5 mostra a exactidão obtida para cada uma das imagens do conjunto de dados de validação. Por exemplo, na classe de imagens, de 83 imagens de campos de arroz, 79 imagens foram classificadas correctamente e 5 não, resultando numa exactidão de 94.0%. Na tabela 5.5, os resultados de exactidão foram calculados considerando todo o conjunto de dados de validação de uma vez.

Classe	Vp	Fp	Fn	TVP	TFN	Exatidão
Arroz	79	4	1	99,0%	1,0%	94,0%
Pereiras	155	1	1	97,0%	3,0%	98,7%
Tomate	33	0	1	97,0%	3,0%	97,1%
Videiras	57	1	0	100,0%	0,0%	98,3%
Total	324	6	3	98,25%	1,75%	97,03%

Tabela 5.5: Resultados dos testes do algoritmo SVM utilizando o conjunto de dados de validação.

Verifica-se que todas as classes de espécies apresentam uma taxa de exactidão muito elevada, com as imagens de validação de tomate e videiras a apresentarem taxa muito perto de (100%). É claro que o modelo não irá prever com essa precisão para essas espécies, sendo possível que para outras imagens de teste classificação seja incorrecta. Em geral, o SVM treinado apresenta um bom desempenho na tarefa de classificação, com uma exatidão global de (97.03%). No entanto, mesmo utilizando dados de validação prevê-se uma diferença nos resultados, pois, durante o processo de validação poderá ter ocorrido *overfitting* incrementando o valor da exactidão, dando uma percepção errada do desempenho do classificador.

5.3 Resultados Gerais dos Testes

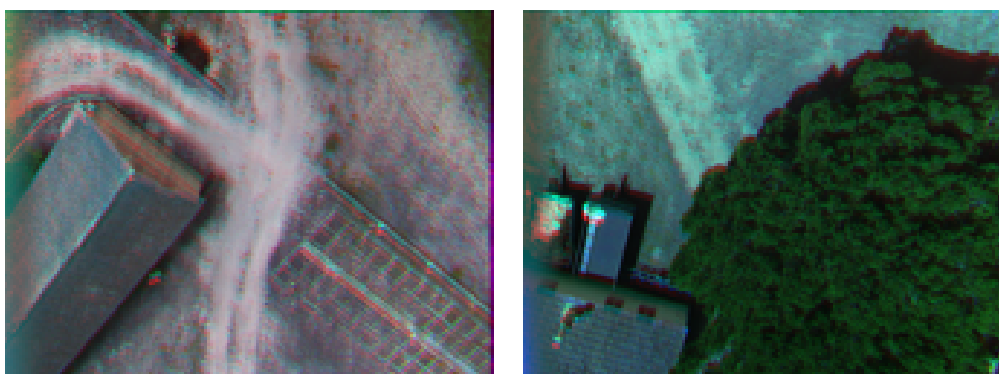
A tabela 5.6 mostra o número de verdadeiros positivos, falsos positivos e falsos negativos resultantes dos testes realizados no conjunto de dados de teste. Como não é possível classificar uma imagem como uma não classe, o número de verdadeiros negativos é zero em todas as classes. No entanto para efeitos gráficos, para traçar a curva ROC, mais especificamente para a taxa de verdadeiro negativo

considerou-se a diferença de imagens entre o conjunto de dados de teste e as classificações atribuídas a essa classe. Ou seja, $Vn = n^{\circ}$ de imagens do conjunto de dados de teste – Vp – Fp – Fn .

Classe	Vp	Fp	Fn	TVP	TFN
Arroz	119	16	7	94,4%	5,6%
Pereiras	236	20	12	95,2%	4,8%
Tomate	47	5	7	87,0%	13,0%
Videiras	72	1	16	81,8%	18,2%

Tabela 5.6: Resultados dos testes do algoritmo SVM utilizando o conjunto de dados de teste.

No caso das imagens que contêm arroz, existem 16 falsos positivos, e como podemos ver na figura 5.6 a maioria deles resultam de imagens que não contêm a classe em questão. Além disso, existem 7 falsos negativos que resultam da classificação de imagens de outras classes como arroz. No caso do teste de imagem das pereiras, existem 12 falsos negativos e 20 falsos positivos. Na classe tomate, existem 7 falsos negativo e 5 falsos positivos. E por fim, a classe videira apresenta 1 falsos positivos e 16 falso negativo. Nesta classe verificou-se que a maioria dos falsos positivos se devia à grande similaridade com a classe pereiras (figura 5.7).



a)

b)

Figura 5.6: Exemplo de imagens onde não está presente nenhuma classe das espécies em questão

Como foi explicado na secção 5.2.1, as imagens que não contem nenhuma classe não foram retiradas pois o objectivo principal era desenvolver o modelo

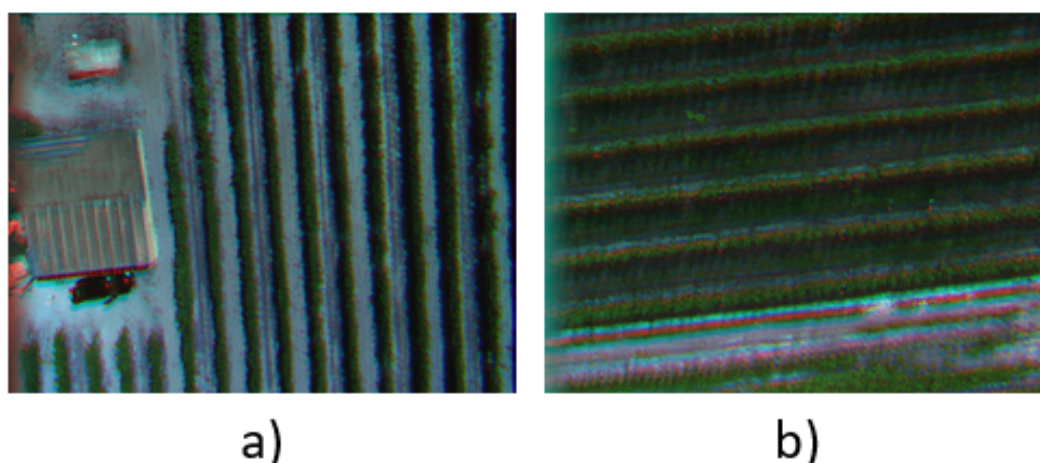


Figura 5.7: Exemplo de imagens onde se verifica a grande semelhança entre espécies e a sua idêntica disposição no terreno. a) Imagem que contém videiras. b) Imagem que contém pereiras.

totalmente autónomo e perceber qual seria a sua eficácia. E por esse motivo, a maior parte dos erros de classificação ocorrem em imagens que não apresentam nenhuma das espécies na imagem. Ainda assim, alguns resultados não possíveis de explicar, pois visualizando os resultados previa-se uma fácil classificação (figura 5.8).

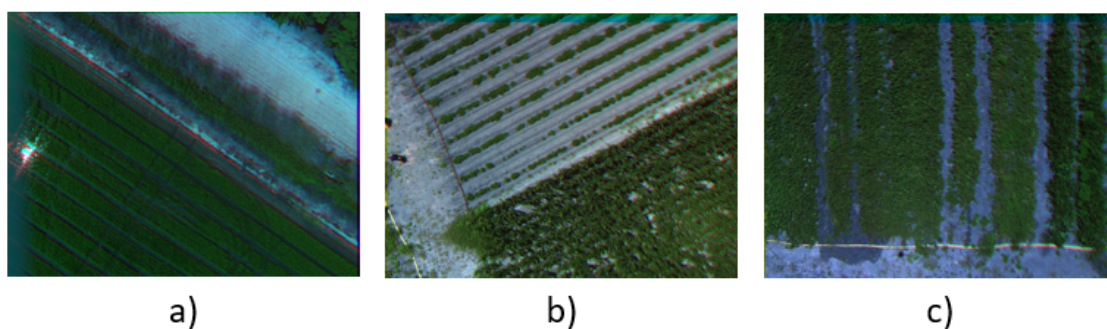


Figura 5.8: Exemplo de imagens em que não foi possível averiguar a razão para a classificação incorrecta. a) Arroz; b) Tomate; c) Tomate.

Esses valores foram aplicados na construção de uma matriz de confusão e de uma curva ROC, representadas nas figuras 5.9 e 5.10, respectivamente. O único propósito desta matriz de confusão e do ROC é mostrar o desempenho do classificador.

A primeira conclusão evidente a se tirar da matriz de confusão é o valor do *recall* diminuiu, aproximadamente em média, 9% em relação ao *recall* dos dados de validação, sendo esta quebra menos acentuada à queda esperada, uma vez que os dados de teste não fizeram parte de nenhum processo de treino ao longo do

		Matriz Confusão					
Classe Verdadeira	Arroz	119	5	2		94.4%	5.6%
	Pereiras	12	236			95.2%	4.8%
	Tomate	4	2	47	1	87.0%	13.0%
	Videiras		13	3	72	81.8%	18.2%
		88.1%	92.2%	90.4%	98.6%	11.9%	7.8%
		Arroz	Pereiras	Tomate	Videiras		
		Classe Prevista					

Figura 5.9: Matriz de confusão resultante dos testes realizados ao conjunto de dados de teste.

desenvolvimento do modelo.

A figura 5.10 possui os mesmos dados, mas representados de uma maneira diferente. Idealmente, toda a curva de cada espécie deve ser semelhante à da figura 5.5b.

No entanto, as curvas que mais se aproximam são as curvas das classes Arroz e Pereiras, pois apresentam um valor de falsos positivos menores que as outras classes. As outras classes de espécies apresentam valores de falsos negativos superiores, fazendo com que a curva fique mais próxima da linha diagonal (linha verde). A linha verde representa a estimativa aleatória. Assim, se alguma das curvas obtidas estivesse abaixo dessa linha, então essa espécie seria classificada de uma forma pior do que a estimativa aleatória. Como esperado, a curva da classe videiras é a mais próxima da linha de estimativa aleatória.

Analisando a tabela 5.6, não é possível concluir muito sobre o desempenho geral do sistema. Assim, combinando as informações da tabela e as equações 5.1, 5.2 e 5.3 foi possível analisar o desempenho do classificador. Os resultados são apresentados na tabela 5.7.

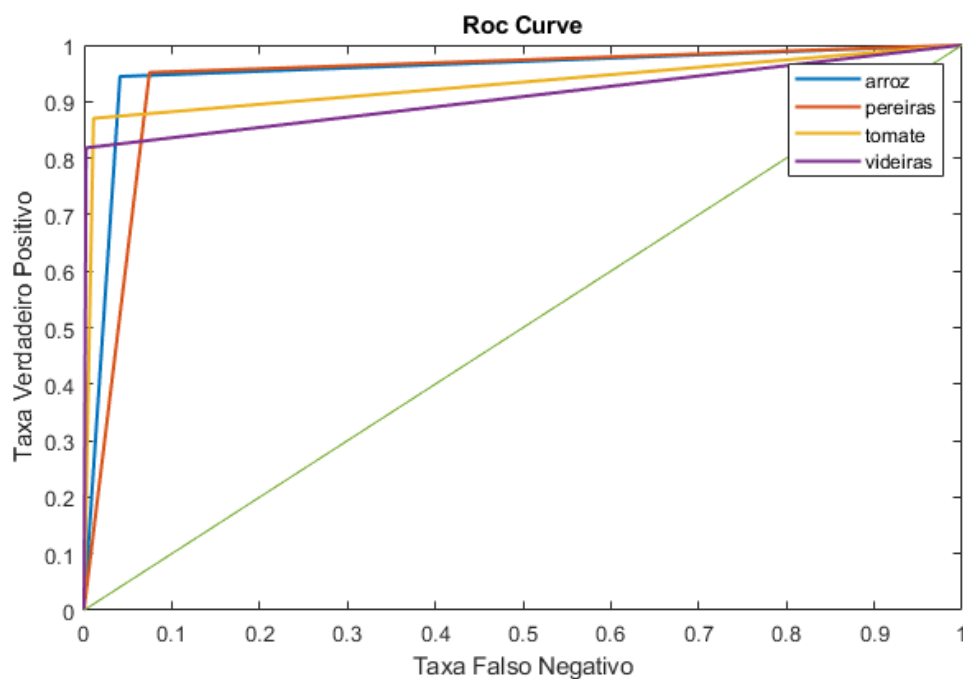


Figura 5.10: Curva ROC resultante dos testes realizados ao conjunto de dados de teste.

Classe	Exatidão	Precisão	<i>Recall</i>
Arroz	83,80%	88,15%	94,44%
Pereiras	88,06%	92,19%	95,16%
Tomate	79,66%	90,38%	87,04%
Videiras	80,90%	98,63%	81,82%
Total	83,11%	92,34%	89,62%

Tabela 5.7: Resultados finais do desempenho geral do sistema com base no conjunto de dados de teste.

No final, os resultados alcançados (figura 5.7) reflectiram um bom desempenho do classificador, já que a maioria dos resultados obtidos superou em média os 80% de sucesso. Fazendo uma análise aos medidores de desempenho apresentados na figura 5.7, é possível concluir que a utilização de RNC é uma ferramenta muito eficaz na extracção de características das imagens, sendo que essas características podem ser utilizadas por qualquer método de classificação de ML. Além disso, foi possível verificar que a própria rede pode ser preparada para classificação de imagens. No entanto, uma vez que não é possível alterar o algoritmo de classificação utilizado pela rede após o treino, esta fica com algumas limitações.

CONCLUSÃO E TRABALHO FUTURO

6.1 Conclusão

De acordo com o objectivo deste projecto, classificação de espécies de vegetação através de imagens aéreas multiespectrais, desenvolveu-se um novo modelo que permite catalogar diversas áreas agrícolas e florestais.

Este projecto está focado na visão computacional, como a extracção das características mais interessantes das imagens adquiridas por uma câmara multiespectral transformando esses dados em informações valiosas, sendo essas informações classificadas por um algoritmo de ML. Os resultados dessa classificação permitem ao usuário identificar diversas culturas de vegetação e a sua localização.

A utilização das RNCs para a extracção das características das imagens, é uma forma fácil de obter mapas de características ricos para a classificação das diversas classes. Nesta dissertação, essas características, utilizadas pelo algoritmo SVM, foram suficientes para distinguir as várias classes de espécies consideradas. Visualmente, as espécies que se previam de mais difícil classificação eram as pereiras e as videiras, devido à sua semelhança. No entanto não se verificaram esses problemas de classificação, apesar de alguns casos de falsos positivos.

Embora a visão global do modelo reflectisse um bom desempenho, os múltiplos métodos utilizados levantaram alguns desafios, como o correto alinhamento e remoção da distorção das imagens, problema que não foi completamente resolvido. A escolha da rede também foi um desafio, pois não foi encontrada nenhuma rede pré-treinada que permitisse a utilização de imagens multiespectrais de 5 canais. Inicialmente não estava planeada nenhuma alteração na estrutura da

rede escolhida, no entanto verificou-se a necessidade de alterar a arquitectura da mesma.

O modelo apresenta muitas vantagens na área da agricultura, permitindo catalogar e identificar as culturas presentes nos campos em estudo. Uma vez que as imagens são georeferenciadas, além da identificação da cultura em estudo é conhecida a localização da mesma através das coordenadas GPS. Esta vantagem pode ser utilizada pelo produtor, tendo o conhecimento da quantidade e localização das culturas que possui, permitindo uma melhor gestão das mesmas. Com a junção de outros modelos desenvolvidos, como por exemplo a saúde das espécies, consegue-se ainda ter o conhecimento do estado das suas culturas. Contudo, o modelo proposto é muito mais vantajoso para fins de fiscalização, uma vez que existe um mercado com um valor substancial pago pelo Estado Português constituído por empresas cuja função é fiscalizar os produtores beneficiados pelos subsídios agrícolas fornecidos pela União Europeia. Essa fiscalização é feita manualmente e pessoalmente, pelos recursos humanos da empresa responsável por fiscalizar, sendo gasto muito tempo e recursos financeiros neste processo. O modelo proposto poderia resolver esse problema. Assim, a utilização de um VANT para recolher os dados seria mais eficiente, pois cobre uma área mais vasta de terreno e o modelo de classificação desenvolvido permite classificar as imagens com elevados índices de sucesso.

O modelo proposto pode ser alargado à floresta. Para isso seriam necessários novos conjuntos de dados que permitissem preparar a rede e o algoritmo de classificação para categorizar as novas espécies. No entanto, na floresta as espécies não estão agrupadas por classes como na agricultura, o que dificultaria a correta classificação das imagens.

A classificação das espécies por imagens apresenta algumas vantagens em relação à classificação das espécies individualmente, sendo a principal vantagem o tempo de preparação dos dados de treino, uma vez que é necessário identificar em todas as imagens de treino cada espécie existente. Como a RNC necessita de um conjunto substancial de imagens esta solução exige demasiado tempo, sendo por esse motivo descartada.

O modelo proposto permitiu obter resultados com uma eficiência global superior a 80% de sucesso, o que nos permite concluir que o modelo proposto é um bom classificador de espécies de vegetação.

6.2 Trabalho Futuro

O objectivo deste projecto cobre uma ampla gama de casos de uso, moldando os mais diversos cenários de aplicação. Embora esta abordagem ao problema da classificação de espécies se destinasse à classificação de imagens, atribuindo apenas uma classe a cada imagem, ainda há muitas questões e desafios que ficam por responder, melhorias dos métodos e novas abordagens de classificação para a bordar no futuro.

Uma vez que as RNCs melhoram o seu desempenho com o aumento do número de dados de treino, no futuro seria interessante estudar a reacção da rede a esse aumento, verificando qual seria o limite da eficiência da rede. Outro ponto que foi focado na secção 5.2.1, e que constitui uma das principais dificuldades foi o alinhamento e a remoção da distorção das imagens. Por esse motivo, seria necessário melhorar o alinhamento e a correcção da distorção das lentes da câmara, ou desenvolver um novo algoritmo com melhor desempenho.

Como apresentado anteriormente, o modelo proposto não permite a classificação individual de cada espécie. No entanto, o passo seguinte passaria por utilizar os ortomosaicos, e ao classificar as imagens que formam esse ortomosaico criar uma grelha que nos indicasse onde se encontravam as espécies classificadas e quais as espécies presentes no ortomosaico. O modelo proposto, com algum processamento adicional, já permite classificar as imagens do ortomosaico, sendo apenas necessário desenvolver a grelha com o resultado da classificação de cada imagem.

Para individualizar as espécies existem outros dados disponíveis, como o *Digital Elevation Model* (DEM), que permite distinguir as espécies pela altura em relação ao solo. No entanto para utilizar o DEM os dados utilizados teriam que ter características diferentes. No caso dos dados florestais, por exemplo espécies de árvores, os valores das alturas envolvidos seriam superiores permitindo a distinção. Contudo, no caso da agricultura, nas alturas das espécies, por exemplo, arroz, os valores são residuais impossibilitando a utilização do DEM. Outras ferramentas como o processamento 3D, como *PointClouds* também permitem a individualização das espécies. O processamento das *PointClouds* permite extrair outro tipo de características que não são perceptíveis no processamento 2D.

Outra abordagem possível ao problema seria o desenvolvimento de um mapa que permitisse identificar as espécies por zonas, ou até mesmo identificar espécies que não são possíveis de individualizar, como por exemplo o arroz e o tomate. Neste caso seria utilizada a segmentação semântica que permite atribuir a cada pixel uma classe criando uma zona onde está presente aquela espécie.

Muitas outras soluções podem ser implementadas e melhoradas, uma vez que o campo da visão computacional e da inteligência artificial, representam um importante tema de atenção no futuro e aspira a um impacto crescente na sociedade. A combinação destas duas áreas tem sido fundamental na resolução de problemas relacionados com saúde e qualidade florestal, prevenção e apoio estratégico no combate a incêndios, e por fim, fiscalização e monitorização florestal e agrícola.

BIBLIOGRAFIA

- [1] E. Adam, O. Mutanga e D. Rugege. “Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation : a review”. Em: (2010), pp. 281–296. DOI: 10.1007/s11273-009-9169-z.
- [2] agronegocios.eu. *Portugal 2020: tudo o que precisa saber para se candidatar*. URL: <http://www.agronegocios.eu/noticias/portugal-2020-tudo-o-que-precisa-saber-para-se-candidatar/>. (acedido: 26.11.2020).
- [3] agrotech. *O futuro da Agricultura de Precisão em Portugal*. URL: <http://www.agrotec.pt/noticias/o-futuro-da-agricultura-de-precisao-em-portugal/>. (acedido: 26.11.2020).
- [4] I. Aicardi, M. Garbarino, A. Lingua, E. Lingua, R. Marzano e M. Piras. “Monitoring Post-Fire Forest Recovery Using Multi-Temporal Digital Surface Models Generated From”. Em: *EARSeL eProceedings* 15.1 (2016), pp. 1–8. DOI: 10.12760/01-2016-1-01.
- [5] A. E. Akay, H. Oğuz, I. R. Karas e K. Aruga. “Using LiDAR technology in forestry activities”. Em: *Environmental Monitoring and Assessment* 151.1-4 (2009), pp. 117–125. ISSN: 01676369. DOI: 10.1007/s10661-008-0254-1.
- [6] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal e V. K. Asari. “The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches”. Em: (2018). arXiv: 1803.01164. URL: <http://arxiv.org/abs/1803.01164>.
- [7] L. Ballanti, L. Blesius, E. Hines e B. Kruse. “Tree species classification using hyperspectral imagery: A comparison of two classifiers”. Em: *Remote Sensing* 8.6 (2016), pp. 1–18. ISSN: 20724292. DOI: 10.3390/rs8060445.
- [8] J. Berni, P. Zarco-Tejada, L. Suárez, V. González-Dugo e E. Fereres. “Remote sensing of vegetation from UAV platforms using lightweight multispectral and thermal imaging sensors”. Em: *Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci* 38 (2009), 6 pp. ISSN: 1380-3743. DOI: 10.1007/s11032-006-9022-5.

- [9] B. Cameras. *SIX BAND CAMERA*. URL: <https://buzzard.camera/six>. (acedido: 09.03.2021).
- [10] A. Canziani, A. Paszke e E. Culurciello. "An Analysis of Deep Neural Network Models for Practical Applications". Em: (2016), pp. 1–7. arXiv: 1605.07678. URL: <http://arxiv.org/abs/1605.07678>.
- [11] A. Carleer e E. Wolff. "Exploitation of Very High Resolution Satellite Data for Tree Species Identification". Em: *Photogrammetric Engineering and Remote Sensing* 70.1 (2004), pp. 135–140. ISSN: 00991112. DOI: 10.14358/PERS.70.1.135.
- [12] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, C. J. Taylor e V. Kumar. "Counting Apples and Oranges with Deep Learning: A Data-Driven Approach". Em: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 781–788. ISSN: 23773766. DOI: 10.1109/LRA.2017.2651944.
- [13] M. L. Clark, D. A. Roberts e D. B. Clark. "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales". Em: 96 (2005), pp. 375–398. DOI: 10.1016/j.rse.2005.03.009.
- [14] O. documentation. *The Kernel Trick in Support Vector Classification*. URL: <https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>. (acedido: 04.11.2020).
- [15] T. Dong, J. Zhang, S. Gao, Y. Shen e J. Fan. "Single-tree detection in high-resolution remote-sensing images based on a cascade neural network". Em: *Canadian Historical Review* 7.9 (2018). ISSN: 17101093. DOI: 10.3390/ijgi7090367.
- [16] K. L. Du e M. N. Swamy. "Neural networks in a softcomputing framework". Em: *Neural Networks in a Softcomputing Framework* (2006), pp. 1–566. DOI: 10.1007/1-84628-303-5.
- [17] J. Everaerts. "The use of unmanned aerial vehicles (UAVs) for remote sensing and mapping". Em: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.March (2008), pp. 1187–1192. ISSN: 16821750.
- [18] J. B. Feret e G. P. Asner. "Tree species discrimination in tropical forests using airborne imaging spectroscopy". Em: *IEEE Transactions on Geoscience and Remote Sensing* 51.1 (2013), pp. 73–84. ISSN: 01962892. DOI: 10.1109/TGRS.2012.2199323.

- [19] A. Fritz, T. Kattenborn e B. Koch. "UAV-BASED PHOTOGRAMMETRIC POINT CLOUDS & TREE STEM MAPPING IN OPEN STANDS IN COMPARISON TO TERRESTRIAL LASER SCANNER POINT CLOUDS". Em: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-1/W2.September (2013), pp. 141–146. ISSN: 1682-1777. DOI: 10.5194/isprsarchives-xl-1-w2-141-2013.
- [20] J. Gertler. "U . S . Unmanned Aerial Systems". Em: (2012).
- [21] R. Gini, D. Passoni, L. Pinto e G. Sona. "Use of unmanned aerial systems for multispectral survey and tree classification: A test in a park area of northern Italy". Em: *European Journal of Remote Sensing* 47.1 (2014), pp. 251–269. ISSN: 22797254. DOI: 10.5721/EuJRS20144716.
- [22] GISGeography. *IKONOS Satellite Imagery: First Commercial Space-based Imaging*. URL: <https://gisgeography.com/ikonos-satellite/>.html. (accedido: 06.02.2020).
- [23] C. Hung, Z. Xu e S. Sukkarieh. "Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV". Em: *Remote Sensing* 6.12 (2014), pp. 12037–12054. ISSN: 20724292. DOI: 10.3390/rs61212037.
- [24] Keras. *Keras Applications*. URL: <https://keras.io/api/applications/>. (accedido: 05.11.2020).
- [25] J. R. K. Lehmann, F. Nieberding, T. Prinz e C. Knoth. "Analysis of unmanned aerial system-based CIR images in forestry-a new perspective to monitor pest infestation levels". Em: *Forests* 6.3 (2015), pp. 594–612. ISSN: 19994907. DOI: 10.3390/f6030594.
- [26] W. Li, H. Fu, L. Yu e A. Cracknell. "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images". Em: *Remote Sensing* 9.1 (2017). ISSN: 20724292. DOI: 10.3390/rs9010022.
- [27] F. Mancini, M. Dubbini, M. Gattelli, F. Stecchi, S. Fabbri e G. Gabbianelli. "Using unmanned aerial vehicles (UAV) for high-resolution reconstruction of topography: The structure from motion approach on coastal environments". Em: *Remote Sensing* 5.12 (2013), pp. 6880–6898. ISSN: 20724292. DOI: 10.3390/rs5126880.
- [28] J. R. Martinez-de Dios, B. C. Arrue, A. Ollero, L. Merino e F. Gómez-Rodríguez. "Computer vision techniques for forest fire perception". Em: *Image and Vision Computing* 26.4 (2008), pp. 550–562. ISSN: 02628856. DOI: 10.1016/j.imavis.2007.07.002.

- [29] A. Matese, P. Toscano, S. F. Di Gennaro, L. Genesio, F. P. Vaccari, J. Primicerio, C. Belli, A. Zaldei, R. Bianconi e B. Gioli. “Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture”. Em: *Remote Sensing* 7.3 (2015), pp. 2971–2990. ISSN: 20724292. DOI: 10.3390/rs70302971.
- [30] Matlab. *trainingOptions*. URL: <https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html>. (acedido: 12.11.2020).
- [31] L. Merino, F. Caballero, J. R. Martínez-De-Dios, I. Maza e A. Ollero. “An unmanned aircraft system for automatic forest fire monitoring and measurement”. Em: *Journal of Intelligent and Robotic Systems: Theory and Applications* 65.1-4 (2012), pp. 533–548. ISSN: 09210296. DOI: 10.1007/s10846-011-9560-x.
- [32] MicaSense. “Multispectral Camera User Manual”. Em: (2017), pp. 1–47.
- [33] R. Näsi, E. Honkavaara, P. Lyytikäinen-Saarenmaa, M. Blomqvist, P. Litkey, T. Hakala, N. Viljanen, T. Kantola, T. Tanhuanpää e M. Holopainen. “Using UAV-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level”. Em: *Remote Sensing* 7.11 (2015), pp. 15467–15493. ISSN: 20724292. DOI: 10.3390/rs71115467.
- [34] S. Natesan, C. Armenakis e U. Vepakomma. “Resnet-based tree species classification using uav images”. Em: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42.2/W13 (2019), pp. 475–481. ISSN: 16821750. DOI: 10.5194/isprs-archives-XLII-2-W13-475-2019.
- [35] O. Nevalainen, E. Honkavaara, S. Tuominen, N. Viljanen, T. Hakala, X. Yu, J. Hyypä, H. Saari, I. Pölönen, N. N. Imai e A. M. Tommaselli. “Individual tree detection and classification with UAV-Based photogrammetric point clouds and hyperspectral imaging”. Em: *Remote Sensing* 9.3 (2017). ISSN: 20724292. DOI: 10.3390/rs9030185.
- [36] O. Nevalainen, E. Honkavaara, S. Tuominen, N. Viljanen, T. Hakala, X. Yu, J. Hyypä, H. Saari, I. Pölönen, N. N. Imai e A. M. Tommaselli. “Individual tree detection and classification with UAV-Based photogrammetric point clouds and hyperspectral imaging”. Em: *Remote Sensing* 9.3 (2017). ISSN: 20724292. DOI: 10.3390/rs9030185.

- [37] G. Pajares. “Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs)”. Em: *Photogrammetric Engineering and Remote Sensing* 81.4 (2015), pp. 281–329. ISSN: 00991112. DOI: 10.14358/PERS.81.4.281.
- [38] S. Puliti, H. O. Ørka, T. Gobakken e E. Næsset. “Inventory of small forest areas using an unmanned aerial system”. Em: *Remote Sensing* 7.8 (2015), pp. 9632–9654. ISSN: 20724292. DOI: 10.3390/rs70809632.
- [39] N. Ranganathan. “Remote Sensing Methods for Forestry Applications – A Survey”. Em: *Ijsart* 3.5 (2017), pp. 394–398.
- [40] M. S. F. on ResearchGate. *Descritores Locais Baseados em Filtros Log-Gabor para Correspondência de Pontos-Chave entre Imagens*. URL: https://www.researchgate.net/figure/Figura-8-Exemplo-de-imagem-multiespectral_fig1_326352343.html. (acedido: 06.02.2020).
- [41] H. Salman, J. Grover e T. Shankar. “Hierarchical Reinforcement Learning for Sequencing Behaviors”. Em: 2733 (2018), pp. 2709–2733. DOI: 10.1162/NECO.arXiv: 1803.01446. URL: <http://arxiv.org/abs/1803.01446>.
- [42] I. Synopsys. *Convolution 2D*. URL: https://embarc.org/embarc_mli/doc/build/html/MLI_kernels/convolution_2d.html. (acedido: 03.11.2020).
- [43] E. O. System. *LANDSAT 5 (TM)*. URL: <https://eos.com/landsat-5-tm/>. (acedido: 06.02.2020).
- [44] M. X. Tagle Casapia, X. Tagle e Ximena. “Study of radiometric variations in Unmanned Aerial Vehicle remote sensing imagery for vegetation mapping”. Em: 23 (2017), p. 60. DOI: 10.13140/RG.2.2.16940.36485. URL: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8918735&fileId=8918738> 0Ahttps://www.mendeley.com/research-papers/study-radiometric-variations-unmanned-aerial-vehicle-remote-sensing-imagery-vegetation-mapping/?utm_source=desktop&utm_medium.
- [45] L. Tang e G. Shao. “Drone remote sensing for forestry research and practices”. Em: *Journal of Forestry Research* 26.4 (2015), pp. 791–797. ISSN: 19930607. DOI: 10.1007/s11676-015-0088-y.
- [46] C. Torresan, A. Berton, F. Carotenuto, S. F. Di Gennaro, B. Gioli, A. Matese, F. Miglietta, C. Vagnoli, A. Zaldei e L. Wallace. “Forestry applications of UAVs in Europe: a review”. Em: *International Journal of Remote Sensing* 38.8-10 (2017), pp. 2427–2447. ISSN: 13665901. DOI: 10.1080/01431161.2016.1252477. URL: <http://dx.doi.org/10.1080/01431161.2016.1252477>.

- [47] P. M. Treitz e P. J. Howarth. "Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems". Em: *Progress in Physical Geography: Earth and Environment* 23.3 (1999), pp. 359–390. ISSN: 0309-1333. DOI: 10.1177/030913339902300303.
- [48] J. Tuominen, T. Lipping, V. Kuosmanen e R. Haapane. "Remote Sensing of Forest Health". Em: *Geoscience and Remote Sensing* (2009), pp. 161–166. DOI: 10.5772/8283.
- [49] D. Wilimitis. *Introduction to Support Vector Machines*. URL: https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. (acedido: 03.11.2020).
- [50] C. W. Woodall, G. M. Domke, D. W. Macfarlane e C. M. Oswalt. "Forests and climate change: forcings, feedbacks, and the climate benefits of forests." Em: *Forestry - Institute of Foresters of Great Britain*. 85.1 (2012), pp. 1444–1449. ISSN: 0015-752X. DOI: 10.1139/X09-086.
- [51] H Zhang e C. Y. Jim. "Species adoption for sustainable forestry in Hong Kong ' s degraded countryside". Em: *International Journal of Sustainable Development & World Ecology* 20.6 (2013), pp. 484–503. ISSN: 1350-4509. DOI: 10.1080/13504509.2013.818590. URL: <http://dx.doi.org/10.1080/13504509.2013.818590>.
- [52] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu e F. Fraundorfer. "Deep learning in remote sensing: a review". Em: december (2017). DOI: 10.1109/MGRS.2017.2762307. arXiv: 1710.03959. URL: <http://arxiv.org/abs/1710.03959> <http://dx.doi.org/10.1109/MGRS.2017.2762307>.